MAY 13 1929

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS
HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XX
No. 5

MAY, 1929

$5.00 a year.
70c per copy.

CONTENTS

THE USE OF QUESTIONS IN SOCIAL SCIENCE MATERIAL. *John N. Washburne*. 321

THE INFLUENCE OF AGE OF PUBESCENCE UPON THE PHYSICAL AND MENTAL STATUS OF NORMAL SCHOOL STUDENTS. *Morris S. Viteles*. 360

ON FORMULA FOR ESTIMATING THE RELIABILITY OF TEST BATTERIES. *Harl R. Douglass and Frederick W. Cozens*. . . 369

A SPECIAL CASE OF THE MULTIPLE CORRELATION COEFFICIENT. *A. K. Kurtz*. 378

A SHORT METHOD FOR COMPUTING THE CORRELATION BETWEEN INTERCHANGEABLE VARIABLES. *Florence L. Goodenough*. . . 386

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 387

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 392

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE APPLETON SERIES IN SUPERVISION & TEACHING

EDITED BY A. S. BARR & WILLIAM H. BURTON

New Volumes

THE NATURE AND DIRECTION OF LEARNING

By *William H. Burton*

All learning and teaching activities related around basic learning types. Material thoroughly tested experimentally. $2.25

THE SUPERVISION OF SECONDARY SUBJECTS

Edited by *Willis L. Uhl*

A compendium of the best practices, by authorities. $2.40

THE SUPERVISION OF ELEMENTARY SUBJECTS

Edited by *William H. Burton*

Experts in all of the subjects on the elementary school level contribute discussions on supervision and improvement of teaching. $2.40

Previously issued

THE ORGANIZATION OF SUPERVISION

By *Fred C. Ayer and A. S. Barr* $2.25

THE SUPERVISION OF INSTRUCTION

By *A. S. Barr and William H. Burton* $2.25

VISITING THE TEACHER AT WORK

By *C. J. Anderson, A. S. Barr, and Maybell G. Bush* $2.00

D. APPLETON AND COMPANY

35 West 32nd St. New York

The Journal of Educational Psychology

is published monthly, except June, July, and August by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in the December number.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single current issues are 60 cents each and less than a full year's subscription is charged at single issue rate for every month ordered. Back volumes $4.50 each; single back issues 75 cents.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which such change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Manuscripts for publication may be submitted to any one of the Editors or to the publishers direct. Books for review should be addressed to the publishers, 10 E. Centre Street, Baltimore, Md.

Warwick & York, Inc.,

Baltimore, Md.

HARVARD UNIVERSITY

Summer School of Arts and Sciences and of Education

JULY 8—AUGUST 17, 1929

The Psychology of Mentally Deficient, Retarded, and Psychopathic Children

Course for School Psychologists and Special Class Teachers

Lectures, clinics, demonstration classes for the study of the methods of training given by Professor Walter F. Dearborn, with the coöperation of Dr. Leonhard Seif, Neurologist and Director of Child Guidance Clinic, Munich, Germany; Dr. Ransom A. Greene, Superintendent of the Walter E. Fernald State School, Waverley, Massachusetts; Dr. Karl M. Bowman, Chief of Staff, Boston Psychopathic Hospital, and of other specialists in this field.

For information concerning this course and other courses in Education and related subjects, address the

Secretary of the Summer School

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS

Manager

HOTEL RENNERT

BALTIMORE, MARYLAND

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX | May, 1929 | Number 5

THE USE OF QUESTIONS IN SOCIAL SCIENCE MATERIAL[1]

JOHN N. WASHBURNE

Teachers College, Syracuse University

The Problem

Is the use of questions in social science reading material a matter of taste, or does it involve a genuine difference in learning? What is the effect of their use upon (*a*) the recall of facts, (*b*) the making of generalizations? Should questions be grouped at the end of the chapter, at the beginning of the chapter, or scattered throughout the chapter? Such are the principal problems the present experiment attempts to solve.

Other questions upon which it throws some light are:

Do questions aid factual recall by attracting attention to one set of facts at the expense of other facts? Do they aid the recall of only those facts with which they deal or is there a spread to other facts?

What is the effect of questions calling for generalizations? Do questions which tend to increase generalizing about a story affect favorably, unfavorably, or not at all the recall of factual material?

Is the effect of the type and placement of questions the same upon children of different grades and ages, boys and girls, the bright and the dull?

Outline of Procedure

I. An historical account of about 3000 words (the story of Florence) unfamiliar to American school children and rich both in factual and generalization material was written in five forms—S, T, Y, Z, X:

[1] This is one of a series of studies made in the Social Science Laboratory of Dr. Harold Rugg, in the Lincoln School of Teachers College, Columbia University. Grateful acknowledgment is made to Dr. Rugg for financing and criticizing the study.

S, with no questions.

T, with questions all at the beginning of the story.

Y, with questions interspersed at the beginning of appropriate paragraphs.

Z, with questions interspersed at the end of appropriate paragraphs.

X, with the questions all at the end of the story.[1]

Some of the questions dealt only with facts which were mentioned in the reading matter; some dealt with generalizations which were not made in the story but could be educed from the facts presented.

II. The five forms of reading material were distributed among equivalent groups of junior high school children.

III. Identical tests were given. The tests had five sections:

1. Factual questions not related to questions in the story. (In the discussion of results, scores in this part of the test are referred to as scores in "non-experimental facts.")

2. Factual questions directly connected with (in most cases, repeating) questions in the story. (Scores in this part of the test are referred to as scores in "experimental facts.")

3. Generalization questions not connected with questions in the story (referred to as "non-experimental generalizations").

4. Generalization questions related to *factual* questions in the story. (The score in this part of the test is referred to as the score in "semi-experimental generalizations.")

5. Generalization questions directly connected with (in most cases, repeating) generalization questions in the story (referred to as "experimental generalizations").

Thus the test was divided into "non-experimental facts," "experimental facts," "non-experimental generalizations," "semi-experimental generalizations," and "experimental generalizations." Examples of each of these five divisions will be given in the detailed account of the experimental procedure.

---

[1] This unalphabetical arrangement of the forms is followed throughout, and it may assist the reader in interpreting the tables and charts to note the pattern of the arrangement. The first is the control group form. In the two principal bar graphs this form is not represented by a bar, but constitutes the line from which the other bars rise or fall. The four experimental group forms are arranged according to the placement of the questions. The form in which the questions most precede the reading matter to which they refer, comes first; the form in which the questions only slightly precede the reading matter to which they refer, comes next; the form in which the questions follow closely the matter to which they refer, comes next; and the form in which the questions follow most remotely the matter to which they refer, comes last.

IV. Tests were sorted into age, grade, and sex groups, and pupils who were under age or over age for their grades were considered separately.

V. In each age-grade-sex group the average score of the children who had read each form was compared with children of the same group who had read the other forms. Thus the average score in Section I of the test of Grade VII girls between the ages of twelve and thirteen who had read the story without questions was compared with the average score in the same section of the test of Grade VII girls between twelve and thirteen who had read the story with questions interspersed, etc. Average total scores were also compared. Grade was compared with grade, girls with boys, and bright children with dull children. The total number of children involved in these comparisons was 1456. But the principal conclusions are based upon a smaller group of 860 children, comprising five equivalent groups of seventh and eighth grade (and a few ninth grade) boys and girls.

Details of Procedure and Materials

The Story.—Form Y of the story (*i.e.*, question preceding appropriate paragraphs) is reproduced hereafter. The other forms were exactly the same except for a different placement of questions.

The following points should be noted in the story: Questions which call for generalizations, call for a type of generalization which is not common either in American schools or American homes. The average seventh and eighth grade child is entirely unfamiliar with such matters. This type of generalization was chosen so as to avoid so far as possible confusion in the child's mind between what the story taught and what is commonly said to be true.

Some of the generalization questions are answered directly and clearly in the paragraphs which follow them; others may be answered only indirectly, by inference. Most of these generalization questions as well as generalizations in the test are modifications of generalizations made by children of Grades VIII and IX of the Lincoln School of Teachers College who were given the story without questions and were told to list all the general truths which one might learn from the story. Three hundred nineteen generalizations were turned in, and these served as a basis for many of the questions in the story and test.

The reading material as given to the children was mimeographed (single spacing) with questions, when interspersed, appearing as center heads. They were indented and separated from the paragraphs

preceding and following by double the white space used in other parts of the material. In short, they stood out clearly as one glanced at the page.

Twenty-five minutes were allowed for reading the story, so that even the slowest readers had time to finish, and the faster readers who usually finished the first reading in from ten to fifteen minutes had ample time to re-read or review the story if they felt so inclined.

Some children laid the story aside and did not reread. Most of the children, however, continued reading the story or turning over the pages until time was called. The reason for allowing so much time for the study of a story which most children could read through in about twelve minutes was that these conditions resemble the conditions usually surrounding the study of a chapter of social science material.

Before passing out the reading material the following directions were given to the children:

> Each one of you will receive a folder like this. Read as rapidly as you can but read carefully because when you are through you will have a test of your memory and understanding. That means that you are to read so as to understand the general story as well as possible, and to remember as many details as possible. You will have only twenty-five minutes to read it through, so waste no time but be sure to read carefully.

The stories were then passed out with the following check against a possible mixup in the forms:

> Everybody in this row should have a big X at the top of the first page—if you haven't, bring back your folder and get another . . . Everybody in this row should have a big Y at the top of the page . . . etc., etc.

After the folders were passed out:

> If there are any words you do not understand, be sure to ask what they mean. Nothing will be taken off for asking the meaning of a word. We want you to know what all the words mean. You may start reading right away.

FORM Y.—HOW THE FLORENTINE PEOPLE STRUGGLED FOR DEMOCRACY

Nearly everybody has heard the word government used but people do not usually think very carefully about what it means. Suppose your class in school was shipwrecked, and landed on an island. Suppose nobody lived on that island before you got there, but that there was plenty of fresh water, and enough fruit so that if it were divided up evenly all of you could keep alive, and if the ground were cultivated and fishes caught there would be enough food to keep you alive for a long time.

In such a case you would probably need a government—in other words, you would need some one to keep order; some one to keep the selfish ones from taking more than their share of the fruit and to make the lazy ones do their share of the

work. If you chose some one to do this ruling, and if you chose others to help him, they would all be parts of the government; and whenever they made rules and made people obey these rules they would be governing.

Would a government be a democracy if one man, whom everybody liked, had all the power?

Of course, there are many kinds of government. If you happened to be the strongest member of your class, or if you had the only gun, you might make the others obey you for a while because they feared you; and some might obey you because they liked the way you governed. But whatever reason the rest had for obeying you, as long as you had all the power and they had none, your government would be called an autocracy.

If, on the other hand, you were at the head of the government but you had no choice at all about how the people should be governed; if the people elected you to run things according to what they wanted, your government would be what is commonly called a democracy. But if one-half of the people desired one kind of a thing and the other half desired the opposite kind of thing, both halves could not have what they wanted. If, in such a case, you let one-half of the people decide what should be done, and let the other half have no say at all in the government, your government would only be about one-half democratic. In a pure democracy, everybody would have an equal say in running the government.

You see that the problem of government is very complicated even under simple and primitive conditions. Imagine how much more difficult your situation would be if you came upon an island that was already inhabited and governed in such a way that you and all of your class had to live in great misery. Before you could begin to decide how to run the government in favor of everybody, you would have to decide how to take it away from the people who ran it only in favor of themselves.

How many years does the present story of Florence cover?

These are just a few of the questions which people who have tried to change governments have always had to answer. Sometimes they have answered them one way and sometimes another. In the city of Florence, Italy, the people answered these questions many times and in many ways. In the three hundred and thirty years between 1200 and 1530 there were many revolutions in Florence and many different kinds of government.

When did the early merchant guilds overthrow the Grandi?

In very early times the city was ruled by a small group of lords whom the Italians called "Grandi." These Grandi taxed the merchants of Florence ten per cent of all the money they took in, and in many other ways governed so that the merchants were discontented and the working people were poor and miserable. (In this story "working people" or "workers" means only those people who worked for wages. It does not mean business men who worked for themselves.) Finally the merchants came together to form guilds. (The guilds of those days were like clubs or unions of business men.) These guilds later united and formed a strong political party. (A political party is a group of people who are all working together to get a certain kind of government or certain kinds of laws.) The party

of the guilds demanded democracy—not for everybody, but only for themselves. They demanded that the business men should have a say in the government. But they could not get control of the city just by asking for it. There were many years of fighting before finally, in the year 1266, they were victorious.

Were the members of the Lesser Guilds workers or men with small businesses?

As soon as the merchants had overthrown the Grandi, they made a law which put the government of Florence entirely under the control of the guilds. Almost all the wealthy Florentine business men belonged to one guild or another; the bankers had a guild of their own, the wool merchants had a powerful guild, and so did the wool dyers. These, and a few others, were called the greater guilds. There were also lesser guilds, such as the butchers' guild, the druggists' guild and various guilds of craftsmen which had no say at all in the government. Only the greater guilds of bankers and merchants and manufacturers ruled the city. It was very much as if, today, the men who are elected to office in the New York Chamber of Commerce by the wealthy business men of New York should become the absolute rulers of all the people in New York City.

But the rich merchants ruled Florence for their own welfare and cared little more for the poorer people than had the Grandi—especially when the workers and the craftsmen (such as carpenters) wanted higher wages for their labor.

Were the Ciompi workers or business men?

There was one large guild or union which had no say in the government. This was the wool-carders' union. The wool-carders were called "Ciompi" (which is pronounced chee-om-pee). These Ciompi worked in the factories of the cloth merchants. It did not take very long for them to see that as far as they were concerned the new government was hardly any better than the old one.

Did having a say in the government have anything to do with getting justice in Florence?

Every time their guild brought a law-suit against one of the clothing guilds, the merchants changed the law so that the Ciompi could not possibly win the case. If the Ciompi wanted to strike, a law was passed against strikers, and so on, until finally they saw that if they wanted justice, and if they wanted to improve their condition as much as the merchants had improved their own, they also would have to have a say in the government.

Did the Ciompi overthrow all of the merchants?

To ask for a part in the government was one thing, and to get it was quite another. The Ciompi had to fight the merchants just as the merchants had had to fight the Grandi, and many of the lesser merchants' and craftsmen's guilds joined the Ciompi to help them take the government away from the greater guilds, just as their ancestors had joined the merchants to help them take the government away from the Grandi. And again the struggle lasted for almost a century. Finally, in 1378, the Ciompi won a great victory. They overthrew the government of the wealthy merchants, and put in their own leader, Michael Lando.

Is a true democracy possible when different classes of people have different interests?

But many of the small merchants of the lesser guilds who had helped the Ciompi, employed workers themselves, and they did not want these workers to run the government any more than they wanted the big merchants to run it. They had made the same mistake as their ancestors had made when they helped the greater guilds: they had thought that the enemy of their enemy must be their friend. Of course, the workers had also made this mistake in taking the small merchants into their party. But as soon as these two classes of people had control of the government they found that they did not want the same kind of rules at all. Michael Lando tried to please both sides. But, of course, whatever pleased one side did not please the other, so Lando had a hard time of it.

Does disagreement affect a party in any way?

Finally the workers turned against Lando, and set up another government which would do what they wanted, and for a while there were two rival governments in Florence. The ordinary run of people (almost everybody except the leaders) who did not know much about politics, became confused and did not know which side to fight on. This is just what the wealthy merchants wanted. They had plenty of money to hire a foreign army—an army of Englishmen who made a business of fighting for whoever would pay them—and while the people were trying to decide which side of the Ciompi party to fight on, the wealthy merchants and their hired army easily entered Florence and threw both the Lando government and the workers' government out of office.

Was Florence better off before just a few rich families got control of the money?

As soon as the merchants had control of the city again, they exiled, hanged, and tortured their enemies, abolished Lando's laws, and even did away with many of the liberties the people had had before the revolution. This new government of wealthy merchants was worse than the earlier one because now the money and governing power had fallen into the hands of only seventeen rich families, and the government was run for the benefit of just as small a group of people as the group of lords who had first run Florence.

The most powerful of these rich families in the new government was the Albizzi. This family took almost entire control of the government and ran it so selfishly that even some of the other rich families objected to the way things were being managed.

Did the working people try to revolt against the Albizzi before Cosimo de' Medici became their leader?

And as for the working people, time after time they tried to start another revolution, but at first were too weak to succeed. Gradually, however, they grew very powerful again and probably would have been able to throw out the Albizzi government if they had not been tricked by a wealthy banker named Cosimo de' Medici (pronounced, Cosimo de Medichee).

Did the people of Florence learn by the mistakes of their ancestors?

Cosimo de' Medici really wanted to get the government away from the Albizzi family so as to run the city himself for the sake of his own business interests. But he told the people that he was their friend and that he wanted to restore the constitution and give them all a vote. He said this, of course, in order to get the help of the people in his fight against his great business rivals—the Albizzi. And again the people foolishly reasoned that since Cosimo de' Medici was fighting against the man whom they hated, therefore Cosimo de' Medici must be their friend.

Which is more dangerous: a clever enemy who pretends to be your friend, or a duller man who is openly your enemy?

As a matter of fact Cosimo was a much more dangerous enemy to them than the Albizzi, because he was much more clever. His idea of a government was not that it should be run to suit seventeen families, but rather to suit just one man—Cosimo de' Medici.

Why did Cosimo de' Medici let the people elect their own public officials?

"But," he thought, "there is no good reason why I should be a target for people to shoot at. The people may as well think I am their friend, even after I have finished with the Albizzi. I will let them have public officials whom they can elect and whom they can blame if things don't please them. But I'll see to it that I manage all the money so that the government can do nothing without my permission. The people cannot understand complicated money matters, so they will never understand my power in the government, and no matter what the government does I shall be safe."

Where were Cosimo de' Medici's banks?

This was very clever reasoning, but it required still greater cleverness to carry out such plans; and this Cosimo had. Of course, he could not have managed it if he had not been tremendously wealthy to begin with. He owned banks all over Europe.

Why did Cosimo lend money to the government?

He started his scheme by lending huge amounts of money to the government. He did this so that the Albizzi and the other rich families who still controlled the government would not dare to interfere with his business schemes. If they did he could make the city pay back all the money it owed him, and this, of course, the government could not afford to do, the money having been spent for wars and all sorts of things.

When Cosimo had tied the government's hands in this way, he did all kinds of tricky things in business. His chief game was lending money to the people who could help him, and then letting them keep it for a long time without paying it back. He let them know that as long as they helped him get from the government the things that he wanted, they could keep the money that they owed him. At first he did not ask very much of them, because he knew that if he did, they might pay the debts and tell him that they would do as they pleased. But when these people had spent all the money and *could* not pay it back, they were entirely

at his mercy, and had to do whatever he said. Whenever he came across a family that had no need of a loan and that therefore he could not control, he ruined that family with the aid of the others who were afraid not to help him.

When only a few men have most of the wealth of a country, is true democracy possible?

Little by little, he got power over (or ruined) all the other wealthy families in Florence except the Albizzi; and in order to get rid of the Albizzi, he needed not only the help of his wealthy dependents but also of the discontented working people of Florence. So he told the workers that he wanted to help them to overthrow the Albizzi in order to gain a democracy.

When did the people drive the Albizzi from power?

In 1444, by means of a revolution, the people drove the Albizzi from power, took all their money, and exiled the entire family.

Do you think the people of Florence would have been wiser if they had followed a principle of government instead of an individual?

After their revolution was successful, they entrusted the making of the new government to the man who they thought had helped them. And Cosimo was clever enough not to anger them while they still were organized and felt their revolutionary power. He made a constitution which seemed quite democratic; that is, he allowed a great many people to vote. But the men whom the people elected could not do much more than argue; they had very little power.

Do you think true democracy is possible when the leaders are not constantly under the observance and criticism of the people?

Most of the real business of the government was done by a council of business men. These were all men who owed Cosimo de' Medici money, and so did very much as he wanted them to do.

Was Cosimo afraid of the people of Florence?

He feared, however, that some of the council members might join together and get the help of the people to put him out of Florence, and take all of his money away from him just as he had done with the Albizzi. So he tried to make the councilmen as afraid of the people of Florence as they were of him.

Do you think the people of Florence should have followed their leaders and public officials loyally and trustfully, or should they have checked up carefully on what they were doing?

He did it in a very simple way: he managed to give every member of the government—councilmen, judges, tax collectors, police captains, and all—a chance to make extra money dishonestly. In other words, he allowed them to steal money that belonged to the people, to take bribes (to accept money for making unjust decisions in law trials, or for passing laws which were not for the public good). If any of the more influential men in the government were too honest to be tempted

into this sort of conduct, Cosimo de' Medici secretly managed to have them removed from office. But most of the public officials were willing enough to become dishonest and get rich by it.

Why did not the council members join together to expose Cosimo?

Cosimo always let these men know that he was aware of their dishonesty, but would not tell on them as long as they did as he wanted. Thus the council members were afraid to tell the people of Florence how they were being secretly ruled by one wealthy man. In fact, among the council members not one was sure that the others were ruled by Cosimo; and each one was so afraid that his own sins would be found out that he was afraid to talk to the others about the matter.

Did the people really have a democracy under Cosimo de' Medici?

After Cosimo de' Medici had ruled the people secretly in this way many years, by means of his great wealth, trickery, and corruption, the public of Florence, finding that they were no better off with their "democracy" than they had been before, began to lose faith in the democratic form of government. As usual, they reasoned very simply: "Now that we are governing ourselves things go just as badly for us as they did before. So what is the use of worrying about the government? Let anyone rule us, so long as we do not starve."

When the people had come to this state of mind, it was not hard, after Cosimo de' Medici's death, for his son, Lorenzo de' Medici, to declare himself Lord of Florence in 1530, and to take away even the show of self-government and the little independence that the Florentine people had been fighting for for hundreds of years. But the people did not revolt, they were no worse off than when they thought they were governing themselves. In fact, they were better off in some ways: Lorenzo gave the people much entertainment: he supported many artists, and gave many public exhibits, carnivals, fairs and parades. Of course, all this was done out of the money the people themselves paid in taxes. But the people, as usual, did not bother about understanding money matters, and so generous did they think their new ruler that they called him "Lorenzo the Magnificent."

The Test.—At the end of twenty-five minutes the stories were collected and the tests were passed out with the following directions:

No one is expected to get a perfect score on this test. It is too hard for that. Even a grown-up probably could not get a perfect score. On the other hand, part of the test is so easy that any one can do it. So do your very best, and finish every page before you turn to the next page. Be sure and answer *every* question. If you are not sure of some, guess.

The following points should be noted in regard to the test.

Questions 1–5 were used as a check on the pupil's understanding of the technique of the test, and whether or not he had actually read the story in the time allotted. They were extremely simple multiple choice questions such as:

What does the story of Florence tell about? (1) Education. . . (2) The Foreign wars of Florence. (3) Religion. . . (4) How the king of Florence gave the people freedom. . . (5) How the people of Florence were tricked by a rich man.

The use of these questions is further discussed under "Method of equating groups and scoring."

Questions 6–18 constitute Element I of the test. They are factual questions not related to questions in the reading matter. The facts covered by these questions are referred to in the discussion of results as *non-experimental facts.*

Questions 18–29 constitute Element II of the test. They are factual questions repeating verbatim or very nearly verbatim the factual questions in the reading material. The facts covered by these questions are referred to in the discussion as *experimental facts.*

Questions 30 and 31 are causal relationship questions repeating questions in the story. They do not constitute a separate test element, but are included in Element V.

Questions 32–42 constitute Element III of the test. These are generalization questions which repeat, or very nearly repeat, the generalization questions in the story. These generalizations are referred to in the discussion as *experimental generalizations.*

Questions 43–47 constitute Element IV of the test. These are generalization questions which refer to the facts covered by the factual questions in the reading matter—and by the questions in Element I of the test. For example, these two factual questions appear in the story: "Did the people of Florence learn by the mistakes of their ancestors?" "Did the people really have a democracy under Cosimo de' Medici?" And the following generalization questions in Element IV of the test refer in another way to the facts covered by the foregoing questions: "Does the story teach that people always learn by the mistakes of their ancestors?" "Does it teach that people may think they have a democracy when they are really being ruled by the power of an individual?"

Such generalizations are referred to in the discussion as *semi-experimental generalizations.*

Element V includes all of Elements III and IV in addition to the two causal relationship Questions 30, 31. In other words, it comprises all experimental and semi-experimental items which are not factual. The chief advantage of this combination is the greater reliability of the score obtained.

Questions 48–55 constitute Element VI of the test. These are generalizations connected with facts which are not covered by any of the questions in the reading matter. These generalizations are referred to as *non-experimental generalizations.*

Element VII comprises all the factual questions in the test, *i.e.*, Elements I and II.

Element VIII comprises all the generalization questions of the test, *i.e.*, Elements V and VI.

Element IX comprises all the other elements in the test, *i.e.*, it is the total test score.

Following Question 55 there were two other pages of questions which had no significance for the present experiment except as time-consumers. They were added chiefly to occupy those fast workers who might bother the other children when they were through. This added material had the effect of a time limit inasmuch as it discouraged going over the test a second time. But it differed from the time limit in that it rewarded rather than penalized the slow, careful worker. Either technique would be satisfactory providing it was the same for all forms.

The tests were stamped with the letter of the form which had been read by the children taking the test; and in order that no mix-up might occur the precaution was again taken of telling the children in each row to be sure that the proper letter was stamped on their tests.[1]

Equating Groups and Scoring Tests

In order to distribute such variables as teacher and group influences as evenly as possible among the various forms, the following procedure was followed in giving out reading matter and tests:

Children of Grade VII were marched into the auditorium and seated from left to right. Children of Grade VIII were then marched in and seated from left to right behind Grade VII (sometimes the order was reversed). Behind these the pupils of Grade IX were seated, also from left to right. Then the forms were passed out from front to back—one form for each row, starting with a different form in each assembly.

Thus the forms were distributed among the classes in a way similar to the following diagram:

[1] Mimeographed copies of the complete test may be obtained from the author on request.

Form S	Form T	Form X	Form Y	Form Z	Form S	Form T	Form X	Form Y	Form Z
Row	Row	Row	Row	Row	Row	Row	Row	Row	Row
1	2	3	4	5	6	7	8	9	10
Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade
VIIA	VIIA	VIIA	VIIA	VIIA	VIIA	VIIA	VIIA	VIIA	VIIA
VIIB	VIIB	VIIB	VIIB	VIIB	VIIB	VIIB	VIIB	VIIB	VIIB
VIIIA	VIIIA	VIIIA	VIIIA	VIIIA	VIIIA	VIIIA	VIIIA	VIIIA	VIIIA
VIIIB	VIIIB	VIIIB	VIIIB	VIIIB	VIIIB	VIIIB	VIIIB	VIIIB	VIIIB
IXA	IXA	IXA	IXA	IXA	IXA	IXA	IXA	IXA	IXA
IXB	IXB	IXB	IXB	IXB	IXB	IXB	IXB	IXB	IXB

Special classes for retarded children etc. were not included. About 300 children were tested on each form. The total was 1456.

The following test papers were not used in computing the results:

1. Those with two or more errors in the first five questions.
2. Those not finished through Question 55.
3. Those with occasional pages not marked.
4. Those not complete because of careless assembling (*i.e.*, with pages missing).

All told, approximately twenty per cent of the tests were not used.

The reason for discarding tests with more than two errors in the first five questions was as follows:

The object of the experiment was to discover what difference, if any, the use of questions made in the learning of those children who actually studied the material. Therefore, the papers of those who either spent the time doing something besides reading the story, and the papers of those who did not understand the technique of answering the test questions, and the papers of those who for any other reason were totally incapable of answering even the simplest questions about the story were extraneous to the central problem. It is true that the comparative standing of the forms would not have been altered by the inclusion of these extraneous factors equally distributed among them, but the per cent of total score by which the forms would have differed from each other would presumably have been less, and the importance of questions to learning would have appeared to be less than it actually was.

The argument for excluding tests not finished through Question 55 and those in which pages were skipped or missing is the same. There was ample time for children who were not especially disabled

and were actually endeavoring to finish the test to get through Question 55. If carelessness or lack of interest resulted in not doing the work these factors were cast aside as obscuring, though not altering, the results.

Those papers which were retained were assorted according to form, sex, age, and grade. The sections of the test were scored separately, two being deducted for each error and omitted answer.

In order to avoid minus scores a constant, ten, was added to each section and remains in the means of the elements as reported in Table I but not in the total score. The true mean is ten less than that reported for Elements I, II, III, IV and VI; thirty less than that reported for the combination Element V; twenty less than that reported for Element VII; and forty less than that reported for Element VIII. As the PE's, V's. etc. are on the basis of the padded score, the deductions have not been made in the table. However, in figuring the per cent of gain as shown in Fig. 1 the proper deductions were made.

Scores were tabulated in the following manner:

FORM Y.—GRADE VII GIRLS

	Under age, 11 and below				Normal age, 12–13					Over age, 14 and above			
Element I.....	Score	Score	Total	Av.	Score	Score	Score	Total	Av.	Score	Score	Total	Av.
Element II....	Score	Score	Total	Av.	Score	Score	Score	Total	Av.	Score	Score	Total	Av.
Element III...	Score	Score	Total	Av.	Score	Score	Score	Total	Av.	Score	Score	Total	Av.
Total.....	Score	Score	Total	Av.	Score	Score	Score	Total	Av.	Score	Score	Total	Av.

In the tabulations it is possible to find any marked difference in reaction to the forms in children of different age, grade, and sex. The final experimental and control groups, however, were made up chiefly of children of normal age of Grades VII and VIII.

In the main experimental and control groups, on which the diagrams and tables reproduced here are based, the number of boys and girls and the ages and grades of these boys and girls is the same for each form. Thus there were five groups of one hundred seventy-two children, equivalent in age, grade, and sex, who finished the test and indicated by their answers to the first five questions that they understood the technique of answering the test questions, and had a rough, general comprehension of the story they had read. Each of these groups was constituted as follows:

Grade	Age	Boys	Girls	Total
VII	11 and under.....	8	1	75
	12–13.....	40	26	
VIII	12 and under.....	10	2	66
	13–14.....	32	22	
IX	13 and under.....	1	5	31
	14–15.....	13	12	
Total.....	.....	104	68	172

The Results

The diagrams which form the basis of our discussion are founded upon the following table:

Table I[1]

Form S

Element	I	II	III	IV	V	VI	VII	VIII	IX	Boys	Girls
M	17.372	16.401	13.256	11.291	36.285	13.314	33.715	49.605	23.471	23.176	23.663
PE_m	.1306	.1317	.1951	.1366	.2886	.1439	.1919	.3826	.5199	.7914	.6865
SD	2.539	2.561	3.794	2.656	5.612	2.798	3.731	7.439	10.109	9.675	10.380
V	14.615	15.615	28.621	23.524	15.466	21.015	11.066	14.996	43.070	41.746	43.866
H	21	23	16	31	10	12	.....	.....	10		
L	9	8	14	20	8	11	.....	.....	13		

Form T

M	17.279	17.302	14.453	11.605	37.814	13.419	34.634	51.250	25.831	27.058	25.029
PE_m	.1248	.1154	.1800	.1326	.2329	.1427	.2113	.3473	.4854	.7706	.6192
SD	2.427	2.244	3.500	2.579	4.529	2.774	4.109	6.752	9.438	9.421	9.362
V	14.046	12.970	24.216	22.223	11.977	20.672	11.864	13.175	36.538	34.818	37.405
H	17	26	30	34	16	16	.....	.....	16		
L	10	5	3	16	4	9	.....	.....	10		

Form Y

M	16.488	17.041	14.581	11.488	37.773	13.267	33.587	50.953	24.669	25.441	24.163
PE_m	.1239	.1247	.1830	.1367	.2809	.1488	.2185	.3739	.5135	.7754	.6856
SD	2.410	2.424	3.558	2.658	5.462	2.893	4.249	7.271	9.984	9.358	10.366
V	14.616	14.225	24.401	23.136	14.460	21.805	12.651	14.270	40.473	36.783	42.899
H	10	28	29	34	15	17	.....	.....	15		
L	16	8	6	14	7	13	.....	.....	13		

Form Z

M	16.634	16.930	14.000	11.488	37.064	13.372	33.506	50.453	24.012	22.544	24.971
PE_m	.1124	.1217	.1887	.1353	.2939	.1515	.2035	.3827	.5332	.8318	.6867
SD	2.186	2.367	3.669	2.631	5.714	2.945	3.956	7.441	10.367	10.179	10.382
V	13.142	13.981	26.207	22.901	15.417	22.023	11.807	14.748	43.175	45.112	41.576
H	6	26	22	38	13	16	.....	.....	12		
L	16	6	12	15	12	15	.....	.....	10		

Form X

M	16.703	16.837	13.750	11.116	36.448	13.081	33.541	49.529	23.128	22.882	23.288
PE_m	.1282	.1204	.2007	.1302	.2851	.1551	.2258	.3639	.5188	.8563	.6500
SD	2.498	2.341	3.903	2.532	5.543	3.016	4.391	7.075	10.087	10.469	9.828
V	14.925	13.904	28.385	22.778	15.207	23.056	13.092	14.285	43.614	45.752	42.202
H	9	21	27	22	6	15	.....	.....	12		
L	9	3	11	24	10	17	.....	.....	12		

[1] The means are comparable as they appear here. For true means, as explained in "Method of Scoring," subtract ten from the reported means of Elements I, II, III, IV, VI; thirty from Element V; twenty from Element VII; and forty from Element VIII.

V. (Pearson's Coefficient of Variation) is derived from the formula

$$V = \frac{100SD}{Mean}$$

"H" indicates "highest scores" and refers to the number of children making the following scores or more: twenty-one in Element I, twenty in Element II, nineteen in Element III, fifteen in Element IV, forty-five in Element V, eighteen in Element VI, and thirty-nine in Element IX.

"L" indicates "lowest scores" and refers to the number of children making the following scores or less: thirteen in Element I, twelve in Element II, seven in Element III, seven in Element IV, twenty-seven in Element V, nine in Element VI, eight in Element IX. The highest and lowest scores were derived by adding approximately 1½SD to the mean scores of Form S.

In interpreting this table the first question to arise is: How important are the differences in score? There is the question both of statistical importance and of educational importance. Let us take up the latter first. It is, of course, impossible to say just how big a change must be in order to be important educationally. Other questions must be considered as well as the amount of change: How difficult is it to bring about this change? How important is the function in which the change is wrought? Obviously if the average life of the children could be increased by five per cent the method by which this could be accomplished would be important, but to increase the ability to recall capital cities by that amount would not be worth much effort. However, most people would grant that an average increase of 30 per cent or 40 per cent would be worth attaining in any subject worth teaching. Figure 1 shows that by the use of questions preceding the story or paragraph we may improve the average score of a group of junior high school children by as much as forty per cent.

The question of statistical importance, or the reliability of the differences, is answered in Fig. 2 where differences are reported in

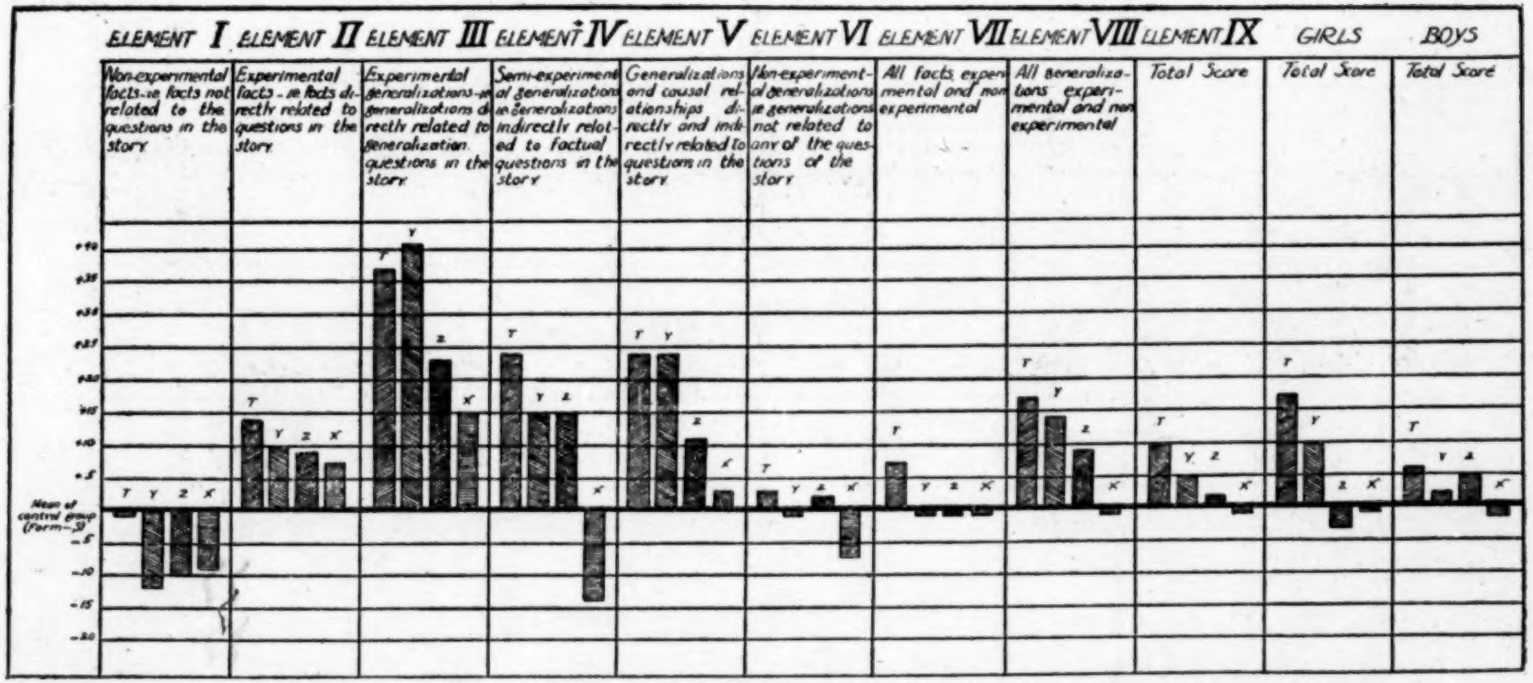

FIG. 1.—Per cents by which mean scores of experimental groups deviate from the control group.

FIG. 2.—Deviations of forms with questions from form without questions.

terms of the probable error of the difference. In this figure as in Fig. 1 the line from which the bars rise and fall represents the mean score made by the control group, *i.e.*, the children who read Form S, the story without questions.

The coefficients of variation are represented in Fig. 2 by the unshaded portions of the bars or unshaded bars with broken lines. The per cent by which the coefficient of variation is reduced is represented by a fall from the line representing Form S. A distance of 1 PE represents a change of 10 per cent in V. The small figures at the base of the unshaded bars indicate the percentage of gain or loss in V.

Inspection of the shaded bars in Fig. 2 shows that in the two "experimental" elements—experimental facts and experimental generalizations—there is a statistically reliable difference between the forms with and the form without questions. The greatest difference throughout is between the form with no questions or between the form with all the questions following the story and the two forms with questions preceding the reading matter—T and Y. Form Z holds an intermediate position throughout.

In every element in which questions in the test are related in any way to questions in the reading (*i.e.*, in all but Elements I and VII) forms T and Y are superior to Forms Z and X.

By and large, T is the superior form as shown in the total score—Element IX. It is reliably superior in several elements and reliably inferior in none.

In Element I (facts not connected with the questions in the reading material) there was a trifling loss of about one per cent or 0.51 PE_{diff}.

In Element II (fact questions directly connected with the questions in the reading) there was a gain of about fourteen per cent in average score (or the statistically important difference in the two means of 5.15 PE_{diff}). In this element it is superior to its nearest competitor, Form Y, by 1.54 PE_{diff}.

In Element III (generalizations directly connected with the questions in the reading) the gain was greater than in any of the other elements. This was true of all the forms. In other words, questions, no matter how placed, affected the making of generalizations far more than the recall of factual material. In Fig. 1 it is evident that those elements involving generalizations (III, IV, V and VIII) show the greatest increase in mean score. Attention should be called to the fact that Element V includes all the questions of Element IV, but as it includes more questions it is more reliable and hence the difference is

statistically more important, although there is little difference in the percentage increase.

In the experimental generalizations (Element III) Form Y appears to be slightly superior to Form T. But as this difference amounts to only one-half of one probable error of the difference (.499 PE_{diff}) it is negligible. The differences in Element IV, V, VI and VIII are also negligible.

In general, then, the two forms in which questions precede the reading may be said to be of about equal merit in everything but factual recall. In this regard, although the superiority of Form T to Form Y in Element II is not statistically or quantitatively very important, when all factual questions are taken together in Element VII, T exceeds Y by 3.45 PE_{diff}.

In total score T also exceeds Y by 1.65 PE_{diff}. But here, as in all elements in which experimental and semi-experimental items are included T does not exceed Y by as much as it does the other forms.

We may conclude from this that what is most important is that the questions shall precede rather than follow the reading material. And that whether the questions precede the entire story or the appropriate paragraph is of less importance. However, the grouping of the important questions at the beginning of the story is probably to be preferred.

It is quite possible, even probable, that many children turned back to these questions after finishing the story and thus had whatever was to be gained by reviewing as well as a certain amount of mind-set to begin with. But this does not explain the superiority of Form T to Form Y, nor of Form Y to Form Z. It is no more likely that children after finishing the story reviewed the questions preceding the story than that they reviewed those preceding or following the paragraphs.

We must account for differences in these forms by differences in mind set. We probably must account for the superiority of all three forms to Form X in the same manner. Although it might be said that in review it would be easier to find the answers to the questions in Forms Y and Z than in X, nevertheless this argument does not apply to Form T. It is just as hard to find answers to questions grouped at the beginning of a story as to find answers to the same questions grouped at the end of the story so far as review is concerned. Thus we may argue that all of the differences existing between T and X, most of the differences existing between all four experimental groups, and a large part of the differences existing between the three

experimental groups—T, Y, and Z—and the control group, are due to differences in mind-set.

If this is true we may conclude that, since the greatest differences in score are in Element III, mind-set is most important in the making of generalizations.

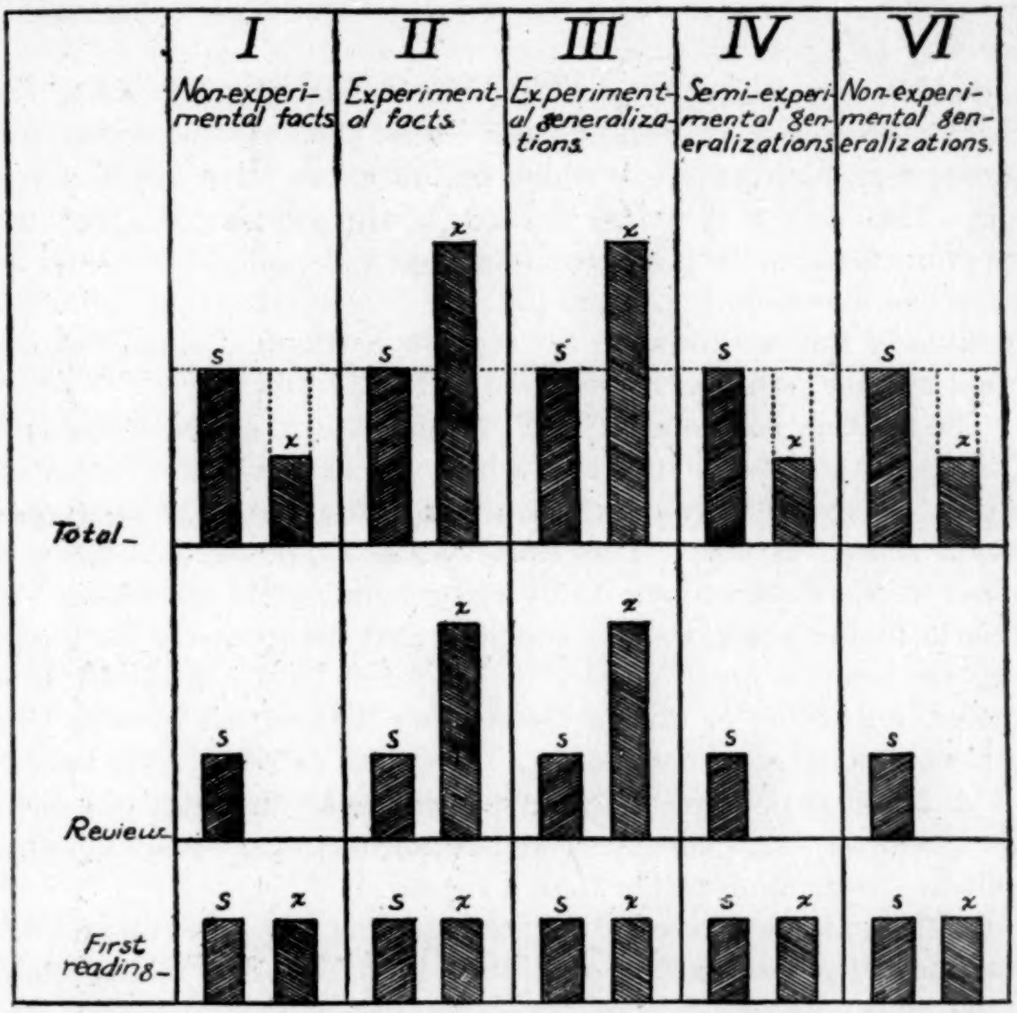

FIG. 3.—Theoretical total scores for Forms S and X showing effect of first reading and review.

So far as the differences between Form X and the control form are concerned, they are presumably entirely due to differences in review. There being no questions in the control form (S), and no questions in Form X until the end of the story is reached, the mind set during the first reading of the two forms is presumably the same. Furthermore, the relative standing of the two forms, S and X, in all elements may be reasonably and adequately accounted for by assuming that the only difference between them is a difference in review.

In Fig. 3 the bottom row of bars represents the theoretical relative standing of S and X in the five basic elements of the test before the readers reached the questions at the end of X. The two forms are represented as being the same in all respects. The middle row represents the theoretical amount by which review raised the scores in the two forms. In this row children reading Form S are represented as spending an equal amount of time reviewing the matter covered by each of the five elements of the test, whereas the children reading Form X are represented as spending the entire time on the experimental facts and generalizations to which the questions have directed attention. The top row is simply the sum of the other two. The dotted line indicates how the graph would appear if the height of S were used as the base line as in Fig. 1 and Fig. 2. It is obvious that the relative positions of the two forms in the top row correspond closely with the actual relative positions represented in Figs. 1 and 2.

The relative positions of T, Y, Z, and X cannot be accounted for in the same manner. In places where X has lost, these forms have gained. They are particularly superior in the matter of semi-experimental generalizations. This element was supposed to indicate the spread of the tendency set up by the generalization questions. If it succeeds in doing this, we may conclude that the spread of the generalizing tendency is much greater in the forms in which questions are encountered before or during the reading than in the form in which questions appear after the reading. Or stated differently, the evidence of Fig. 1 is that preview and interspersed questions result in a noticeable spread of the tendency to generalize, but that review questions result in a restriction rather than a spread.

In Element V we have the results regarding experimental and semi-experimental generalizations and causal items, *i.e.*, everything relating to the questions (directly or indirectly) except factual recall. Here we see the distinct superiority of the preview and interspersed preview forms over the review and interspersed review forms. The review form (X) is not at all reliably superior to the control form (S), revealing here, as it does consistently, its comparative weakness in increasing general comprehension of the story.

Element VI reveals that the review form is the only one that differs at all significantly from the control form in the matter of non-experimental generalizations. There is evidence of loss in the review form (X) here—as in all non- and semi-experimental factors. The preview form (T) shows here, as in Element I, that the great gain it makes in

experimental and semi-experimental factors is not at the expense of the non-experimental factors.

For the two forms in which questions were interspersed Element VI shows no such loss as in Element I, *i.e.*, although their gain in experimental and semi-experimental items resulted in a loss in the recall of non-experimental facts it did not result in a similar loss in the making of non-experimental generalizations. Element VII indicates, as will be shown, that only the interspersed factual questions were responsible for the loss in Element I and that the generalization questions resulted in gain with no compensatory loss.

Element VII is the sum of the scores in experimental and non-experimental facts, these two items being equally weighted. If what is gained in one factor (*e.g.*, experimental facts) is lost in the other (*e.g.*, non-experimental facts) then the sum of the two should be the same for all five forms. This turns out to be the case except for form T.

How may we account for the outstanding superiority of form T in this element? Why should T alone gain in Element II (experimental facts) without losing proportionately in Element I (non-experimental facts)?

The consideration of what patterns or parts of patterns the facts form offers an explanation. All the facts remembered have some significance; they form a part of some pattern. It seems reasonable to suppose that the facts forming one perceived pattern are more easily remembered than the same number of facts forming many patterns. And it also seems probable that the children reading Form T were (during the first reading at least) more likely to fit the facts they encountered into a comprehensive scheme than the children reading the other forms. The questions preceding the story, were arranged in the same sequence as those which were interspersed; thus they foreshadowed the development of the story, and established a point of view. The latter point is important. Every time the point of view shifts groups of facts arrange themselves into new patterns; *i.e.*, they gain a different significance. Thus there were many significances given by the Lincoln School children to the fact (or facts) of Cosimo de' Medici's trickery: Among others:

(1) There is danger in letting one man have all the money.
(2) Be sly and clever when people are doubtful of you.
(3) Don't believe the people in office are going to do what they say they are.
(4) Don't take bribes.
(5) Poorer classes should not be tricked by politicians.
(6) Don't let one man or one faction get too much power.

It is plain to be seen that fewer facts of the story fit into (4) of the foregoing generalizations than into (6). The child who fits the facts concerning graft into a lesson for public officials instead of into the general lesson of the story, which is a lesson for the people who elect the officials, must fit the other facts of the story into some other picture or ignore them. For example, one may imagine this child saying as he reads along without any consistent point of view:

> The Grandi lost out because they taxed the merchants so much; that shows *it does not pay to overtax the people.*[1]
>
> There were merchants and Grandi and workers and so forth; that shows that *there are different classes of people.* There were lots of different rulers in Florence; that shows *there will always be a change in the so-called president.* The Ciompi formed a guild in order to help themselves, but it did not succeed; that shows that *it does not do much good to form clubs against a government.* Cosimo de' Medici succeeded; that shows that trickery[2] pays and the officials under him could not do as they wanted because they took bribes; and that shows[3] that you should not take bribes etc., etc.

When this child has finished the story he has, because of his shifting point of view, many facts put together in many more or less disconnected pictures, whereas the child whose point of view has remained more constant has the same number of facts in a more coherent and uniform design, and can therefore recall them more readily.

In the case of the five forms of the story of Florence, what probably happened is that the children who read Form S, approached it without any definite mind-set except to understand and remember as much as possible, and they reviewed as much of it as they could or would in a state of mind altered only by some knowledge of the scope of the story. The children who read Form X did so originally with the same lack of directed mind-set as those who read Form S, but their review was directed.

Forms Y and Z were read with directed mind-sets. But the directions were given piecemeal and with such specificity that all but those facts essential to answering the questions were excluded from the picture. One might compare the effect of interspersed questions with the effect of focussing a camera sharply upon a nearby object. Everything that is not within a very few feet of the object focussed upon is cut out of the picture. The effect of lumping the questions at the beginning of the story might be compared to the effect of focus-

[1] Italics indicate actual generalizations made by Lincoln School children.

[2] See (2) above.

[3] See (4) above.

sing on a more distant object; more surrounding details and background are brought into the picture.

To speak in different terms the children reading "Form T could not have remembered specifically all of the twenty-four questions which preceded the reading matter. But they probably sensed the type of question which was deemed important, *the point of view* which these questions assumed, and something of the general drift of the story.

It is not unlikely that the prime virtue of Form T was that it established a point of view at the outset.

Another strong point probably was that the child's attention was not too specifically directed once he was set going in the right direction. This may account for the much better standing of this form in non-experimental factors as well as in experimental factors, and, if so, the superiority of this procedure is indicated with especial force in Element IX—the total score. One-half of the test items constituting this total, comprise experimental facts and generalizations, and one-half comprise non-experimental and semi-experimental facts and generalizations.

If this element (IX) is taken to indicate general mastery of the story, the superiority of Form T is clear. It raises the average total score ten per cent which is twice as far as Y raises it. It lowers the variability (V) by 15 per cent, decreases the number of children getting the lowest scores by thirty per cent, and increases the number of children getting the highest scores by sixty per cent. Its total average exceeds that of Y by 1.63 PE_{diff} and that of all the other forms by more than 3.0 PE_{diff}.

We may go back now to a consideration of Element VIII. This element comprises all generalizations—experimental and non-experimental. The same arguments apply here as in Element VII. But the conclusions are different. It appears that here, in the matter of generalizations, gain in experimental factors, in Y and Z, does not involve a corresponding loss in non-experimental factors.

Inspection of the total scores for girls and boys reveals the fact that girls are far more affected by the preview questions (by mind-set?) than are boys.

This finding is interesting from two points of view. As a light on the nature of the differences between the sexes in the junior high school it needs to be considered in connection with age and grade differences and in connection with the responses in the different elements of the

test. Although we have the necessary data the problem appears to be altogether too intricate for adequate discussion in the present study. Suffice it to say that T seems to hold its own for boys and girls of all ages between ten and fifteen. But the greatest differences occur in the scores of girls.

From another point of view, it is interesting to note that had it been possible to include more girls in the equated groups many of our differences—especially gains—would have been increased—by as much as seventy per cent in some instances. As it is, the boys outnumber the girls in our equated groups in the ratio of about three to two.

Effect of Questions upon Correlations of Test Elements

What is the relation of factual recall to ability to generalize? How do the factual questions in the story affect generalizing? How do the generalization questions affect factual recall? How do the experimental and non-experimental factors vary in relation to each other?

These questions are partly answered by the comparison of mean scores in Fig. 1. Some further light may be gained by a consideration of the intercorrelations and partial correlations in Table II.

In general, the correlations are shown to be low and positive. The average of the intercorrelations for Forms S, T, Y, X is .35, and for form Z it is .41.

The highest correlations for most of the forms are between Elements I and II (factual recall). The lowest correlations for the form without questions are between the elements involving factual recall and generalizations—Elements I and III, I and VI, II and III, II and IV, and II and VI. The lowest correlations for forms with questions are between experimental elements and non-experimental generalizations (II and VI, III and VI), and between non-experimental facts and generalizations. In other words, the experimental groups show the same low correlation as the control group between facts and generalizations except where the experimental questions are concerned.

Carrying the comparison between the control and experimental groups further: correlations were most *increased* (by the questions) in Elements II and IV—experimental facts and the generalizations (semi-experimental) connected with those facts; and were most decreased in Elements III and VI—experimental and non-experimental generalizations.

TABLE II.—INTERCORRELATIONS AND PARTIAL CORRELATIONS[1]

Elements	T	$\frac{T - S, \text{Diff.}}{PE_{diff}}$	Y	$\frac{Y - S, \text{Diff.}}{PE_{diff}}$	Z	$\frac{Z - S, \text{Diff.}}{PE_{diff}}$	X	$\frac{X - S, \text{Diff.}}{PE_{diff}}$	S
1 and 2	.561	.85	.501	.10	.481	− .18	.654	2.44	.494
1 and 3	.298	− .08	.408	1.22	.443	1.66	.358	.61	.306
1 and 4	.340	− .41	.344	− .40	.500	1.57	.363	− .21	.379
1 and 6	.300	.03	.247	− .58	.306	.11	.357	.75	.296
2 and 3	.335	− .18	.367	.17	.379	.32	.327	− .32	.353
2 and 4	.360	1.22	.325	1.01	.484	2.94	.275	.43	.238
2 and 6	.272	.48	.271	.54	.341	1.34	.291	.77	.224
3 and 4	.374	− .26	.405	.09	.455	.74	.332	− .84	.399
3 and 6	.269	−1.17	.291	−1.14	.276	−1.33	.208	−2.12	.384
4 and 6	.393	− .62	.473	.24	.478	.31	.292	−2.11	.456
12.3	.52	.....	.41	.....	.37	.....	.61	.....	.42
13.2	.14	.....	.28	.....	.31	.....	.20	.....	.17
14.3	.26	.....	.20	.....	.38	.....	.28	.....	.30
16.3	.24	.....	.15	.....	.20	.....	.31	.....	.20
23.1	.22	.....	.21	.....	.21	.....	.12	.....	.24
24.1	.22	.....	.20	.....	.31	.....	.06	.....	.06
24.3	.27	.....	.22	.....	.37	.....	.19	.....	.12
26.1	.13	.....	.18	.....	.23	.....	.08	.....	.08
26.3	.20	.....	.18	.....	.26	.....	.24	.....	.10
34.1	.30	.....	.31	.....	.31	.....	.23	.....	.32
34.2	.28	.....	.33	.....	.34	.....	.26	.....	.34
34.6	.30	.....	.32	.....	.39	.....	.29	.....	.28
36.1	.20	.....	.24	.....	.16	.....	.09	.....	.32
46.3	.33	.....	.40	.....	.41	.....	.28	.....	.36
64.1	.33	.....	.43	.....	.40	.....	.19	.....	.40
14.32	.15	.....	.12	.....	.28	.....	.21	.....	.28
16.32	.17	.....	.09	.....	.11	.....	.21	.....	.18
46.32	.29	.....	.37	.....	.34	.....	.24	.....	.36
34.12	.26	.....	.28	.....	.26	.....	.22	.....	.32
36.12	.18	.....	.21	.....	.12	.....	.08	.....	.31
64.12	.31	.....	.41	.....	.36	.....	.18	.....	.41
24.13	.16	.....	.14	.....	.26	.....	.04	.....	− .02
26.13	.09	.....	.14	.....	.21	.....	.07	.....	.003
64.13	.29	.....	.40	.....	.37	.....	.16	.....	.33
34.126	.22	.....	.21	.....	.23	.....	.22	.....	.22
24.136	.14	.....	.09	.....	.20	.....	.03	.....	− .02
64.123	.27	.....	.37	.....	.34	.....	.16	.....	.34
14.236	.11	.....	.10	.....	.26	.....	.18	.....	.24

R total score = .88.

[1] Reliability of the total test was estimated from the formula $rx = \frac{2rh}{1 + rh}$ as follows: The intercorrelations of the control form (S) were used; the correlation between factual items and factual items (Elements I and II) was taken and the correlation between generalization items and generalization items (Elements IV and VI) was also taken; these two correlations were averaged, and the average treated as the correlation of one-half of the test with the other. The reason Elements IV and VI were used instead of, *e.g.*, III and VI was those they were more nearly of the same size. The more general formula $rx = \frac{Nr}{1 + (N-1)r}$ including all the intercorrelations was not used because, obviously, the sections of the test with factual questions are not equivalent to sections with generalization questions, and hence the correlation between them does not represent reliability.

The latter fact may be explained on the basis that improvement in experimental generalizations was not accompanied by a corresponding change in non-experimental generalizations. On the other hand, the increase in correlation in Elements II and IV may be explained in forms T, Y, Z on the basis that improvement in experimental facts *was* accompanied by a corresponding change (improvement) in semi-experimental generalizations. It was also accompanied by a corresponding change (significant when Elements I and III are partialed out) in non-experimental generalizations.

These findings point to the strange conclusion that although the correlation between ability to recall facts and ability to make generalizations is lower than the correlation between ability to make one type of generalizations and ability to make another type of generalizations, nevertheless, except for generalizations directly connected with generalization questions, fact questions exert more influence than generalization questions upon the making of generalizations.

Before accepting such a conclusion, however, it is necessary to consider the changes in correlations in connection with the changes in mean scores. These relationships are shown in Fig. 4.

This diagram is based upon the intercorrelations. The changes indicated by the partial correlations are in no instance contradictory to the changes indicated by the intercorrelations. In most instances although the sizes of the correlations are reduced the differences between them are increased. In one case (Elements III and IV) the differences are reduced. This would not, however, materially affect Fig. 4. The value of the partial correlations is not certain, for although the relations measured appear to be linear, no adequate test of linearity was applied.

Figure 4 is constructed as follows: The squares represent scatter diagrams. The diagonal line in each square represents the line of perfect correlation. The arrows represent the direction in which the majority of scores appear to have moved as a result of questions in the reading matter. As usual in scatter diagrams, the low scores of the function measured along the ordinate (Function A) are understood to be at the bottom and the high scores at the top; and the low scores of the function measured along the abscissa (Function B) are understood to be at the left and the high scores at the right.

Thus the lower left-hand corner of each square represents low scores in both functions, and the upper right-hand corner represents high scores in both functions. The upper-left hand corner represents high scores

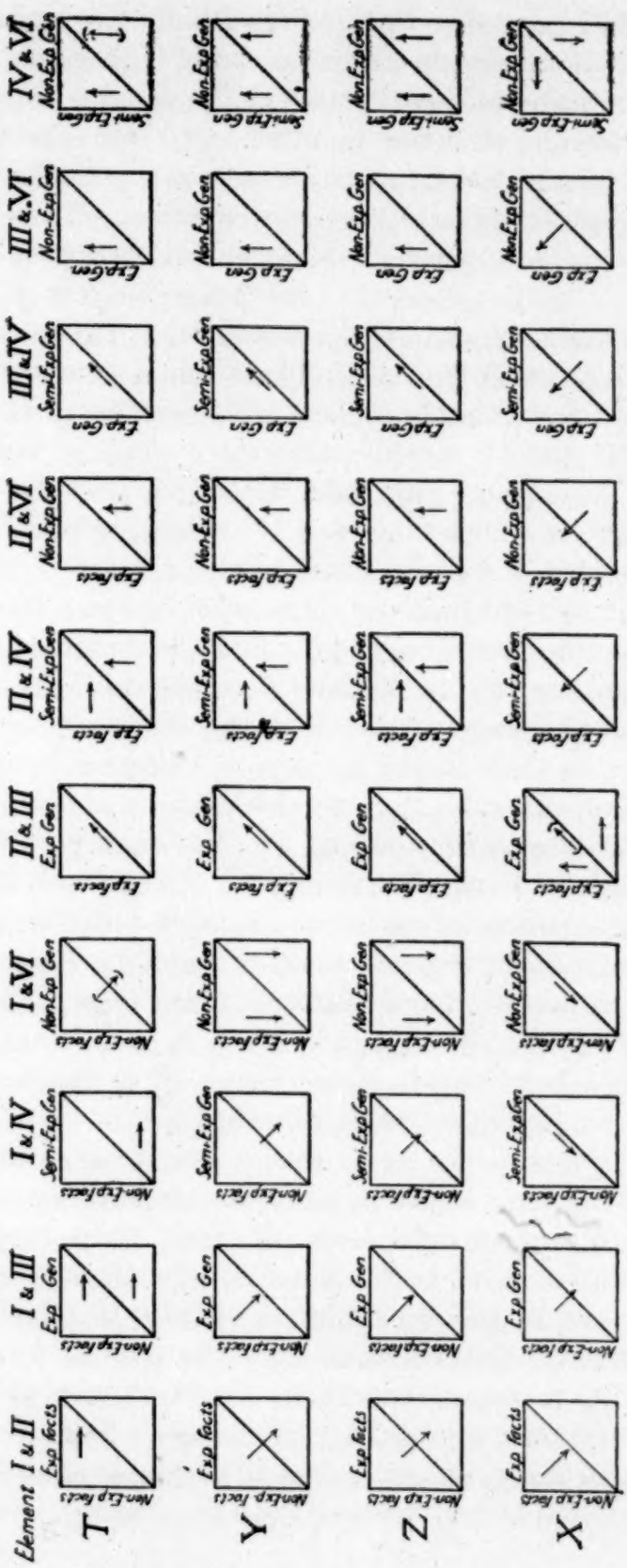

FIG. 4.—Relative change in the five fundamental elements of the test.

in Function A and low scores in Function B, and the lower right hand corner represents low scores in Function A and high scores in Function B. An arrow pointing in the direction of the diagonal line indicates a rise in correlation, and an arrow pointing away from the diagonal line represents a fall in correlation; an arrow bisected by the line or paralleling it indicates practically no change in correlation. Thus a horizontal arrow pointing to the upper right-hand corner, as in the first square under I and II (Fig. 4) means that the average score of Function B (Element II) is higher than in the control group, the average score of Function A (Element I) is practically the same, and the correlation between the two functions is higher. The arrows in the first three squares under III and IV indicate that the average of both functions rose while the correlation remained the same, and the last square under III and IV, in which the arrow is bisected by the line of correlation, indicates that there was a rise in the average of Element III, a fall in Element IV, and that the correlation between them remained the same. Where an arrow crosses the line a little but not enough for the projection on one side to balance the projection on the other, it indicates a slight change of correlation. If more of the arrow approaches the line than leaves it, there is a slight rise, etc. Arrows in parentheses indicate subsidiary tendencies.

The most striking feature of Fig. 4 is the close parallelism in the changes produced by Forms T, Y, and Z. In general, the direction of change in these three forms is the same, but the amount differs. There is an exception in the relationship involving Element I where Form T makes a more independent gain than that made by Forms Y and Z. The direction of change in Form X is different or opposite from that of the other forms in eight of the ten comparisons.

Figure 4 may be interpreted as follows:

Elements I and II.—Form T shows an arrow going from the corner representing high scores in non-experimental facts (Element I) and low scores in experimental facts (Element II) toward the corner representing high scores in both elements. In other words, Form T shows fewer scores in the upper left and more in the upper right corner of the scatter diagram than does the control form (S). We may interpret this to mean that Form T affects most those students who, without questions to guide them, recall some facts well but whose performance is inconsistent. It raises their performance where they are weak without impairing it where they are strong.

The effect of the questions in the other forms is the same in every respect except that there is a loss in one element corresponding to the gain in the other, and in Y and Z those children who without questions would have done moderately and about equally well in both elements are affected as much as those whose scores show discrepancies; *i.e.*, scores in the middle of the scatter diagram are moved down toward the lower right corner as much as those of the upper left corner are moved toward the middle. In Form X, however, there are fewer children who make high scores in non-experimental facts, and low scores in experimental facts, and there are no more children than in Form S who show the reverse discrepency. Therefore there is a rise in correlation.

Elements I and III.—For Elements I and III Form T shows two arrows pointing in the same direction, one approaching the line and the other leaving the line, thus indicating no change in correlation. Those who did well and those who did poorly in Element I are equally affected. Their scores are improved in Element III (generalizations) without loss in Element I. Y and Z show a loss in Element I, and those who did poorly in Element I are less affected than those who did well, and this results in a higher correlation. Form X shows a tendency similar to Y and Z.

Elements I and IV.—Here the situation is about the same as for I and III except that the effect of Form T is that fewer children do poorly in both functions, but more do well in semi-experimental generalizations and poorly in non-experimental facts. Y and Z show results in this comparison similar to the results in Elements I and III.

Form X loses about equally in both factors. Since this form is altered only by review, we are led to the interpretation that both factors—non-experimental facts and semi-experimental generalizations—were neglected in review.

Elements I and VI.—The change that occurs in these two non-experimental factors is so slight in Form T that the scores may best be said to be unaffected. Such trifling change as occurred was in the direction of raising the score in non-experimental generalizations while lowering it in non-experimental facts.

The arrows in Y and Z indicate that the loss in non-experimental facts was not accompanied by a corresponding change in non-experimental generalizations. Form X loses about equally in both factors.

The argument is the same as in the comparison for Elements I and IV.

Elements II and III.—In these two experimental factors all the forms have a similar effect: Both factors are raised, and raised proportionately. The correlation between the experimental facts and the experimental generalizations in the five forms differ by less than one-third of one probable error of the difference. All show a slight tendency to move up the generalization scores more than the fact scores. But the ratio between their means is fairly constant. In Forms S, T, Y, Z, X the mean of Element III is approximately eighty-one per cent, eighty-two per cent, eighty-five per cent, eighty-three per cent and eighty-two per cent respectively of the mean of Element II.

Elements II and IV.—In Forms T, Y, and Z there is a marked rise in correlation between Elements II and IV, and a rise in the mean score of each. This is indicated by the arrows pointing toward the line of correlation from two directions. In other words, both scores are raised and discrepancies in the scores are reduced, *i.e.*, fewer people get a high score in one while getting a low score in the other. Also Table I shows that fewer children made a low score (in Forms T, Y, Z) in both elements and more children made a high score in both elements. An arrow in parenthesis rising along the line of correlation might be included to show this subsidiary tendency. But throughout the diagram (with the exception of T I and VI) subsidiary tendencies are not indicated unless they are quite marked in connection with either the partial or intercorrelations. That Elements II and IV are not affected independently is suggested by the fact that the ratios of their means are scarcely changed in spite of the other marked changes. Thus in Form S the mean of Element IV is twenty per cent of the mean of Element II, and in Forms T, Y, and Z it is twenty-two per cent, twenty-one per cent and twenty-one per cent respectively. This indicates that perhaps the two arrows should be represented as pointing diagonally toward the line of correlation from opposite corners, instead of pointing vertically and horizontally as in the diagram. Such a representation would suggest the interesting interpretation that gain in either factor was at the expense of the other, and that neither factor was studied to the exclusion of the other. This could be explained on the hypothesis that the factual questions caused a tendency in one direction and the generalization questions caused a tendency in the opposite direction.

Form X shows a loss in one of the two factors proportional to the gain in the other. Fewer children do well in semi-experimental generalizations and poorly in experimental facts. But the correlation between the two factors remains about the same. The fact that one factor gains at the expense of the other looks again as if, in review, the facts had been considered in isolation and their significance had not been seen.

Elements II and VI.—In Form X the relative change in Elements II and VI is the same as in II and IV. In the other forms the change is, as it were, half of what it is in II and IV; *i.e.*, discrepancies of only one sort, instead of two, tend to disappear. Fewer children get high scores in non-experimental generalizations and low scores in experimental facts. But just as many get high scores in the latter and low scores in the former. Another way of interpreting the straight rise of the arrow is that the scores of the better students (those good enough to get a high score in non-experimental generalizations) were improved in Element II without affecting their standing in Element VI.

Elements III and IV.—The relative change wrought by Form X is the same in Elements III and IV as in the preceding two comparisons except that the correlation remains the same instead of being slightly raised. (The arrow is bisected by the line of correlation instead of being crossed just below the head.) The other forms show a gain in both factors without either a reduction or an increase in the number of children scoring high in one factor and low in the other.

Elements III and VI.—Form X shows a change in III and VI somewhat similar to the changes in the three preceding comparisons, but with a marked lowering of correlation (2.12 PE_{diff} lower than the correlation in Form S) due to a larger number of children making low scores in non-experimental generalizations and high scores in experimental generalizations.

In the other three forms also there was a decided increase in the number of children scoring high in Element III and low in Element VI. But the mean of VI was not lowered as in Form X, so the rise in III seems to be independent of a change in VI. A subsidiary tendency in the scores of Form T is toward a rise in the non-experimental generalizations corresponding to the rise in experimental generalizations. There are fewer low scores in both factors and more high scores, and the means of both rise. There are likewise more

high scores in non-experimental generalizations for Forms Y and Z, but these are counterbalanced by more low scores.

Elements IV and VI.—Form X seems to have resulted in fewer children scoring well in both non-experimental and semi-experimental generalizations, but in no reduction of the number of children scoring high in one element while scoring low in the other. This results in a correlation between the two elements lower by over 2 PE $_{diff}$ than the correlation in the other forms.

Forms T, Y and Z result in fewer children doing well in non-experimental generalizations and poorly in semi-experimental generalizations, but in more children doing poorly in non-experimental generalizations and well in semi-experimental generalizations. Thus the correlation remains practically the same as in the control group, while the mean of one element is raised without a corresponding change taking place in the mean of the other.

Figure 4 has shown that the increase in correlation between Elements II and IV should not lead to the conclusion, suggested by considering the correlations without reference to the mean scores, that "fact questions exert more influence than generalization questions upon the making of generalizations." The contrary is the case. The change in III and IV is more closely related than the change in II and IV, although the latter shows greater increase in correlation.

The third order partials were intended to throw further light on this problem. But they are difficult to interpret. The correlations between III and IV are consistently higher than between II and IV. But they are still higher, in most instances, between VI and IV. The results might be quite different if the influence of intelligence could be partialed out. But although this factor is held constant as between forms, its influence upon the relation of one element of the test to another is unknown. We must therefore rely for our interpretation upon Fig. 4, and use our partials merely as a check on the intercorrelations in the construction of this diagram.

Conclusions

In the statement of our problem we listed a number of questions which our procedure was designed to answer. Let us see now what answers have been found.

1. *Is the use of questions in social science reading material a matter of taste, or does it involve a genuine difference in learning?*

A genuine difference in learning amounting in the case of experimental generalizations to 40 per cent of the mean score or over 5 PE_{diff} is involved in the use of questions.

2. *Is the placement of questions in social science material a matter of taste or does it make a genuine difference in learning?*

If the two extreme Forms T and X (all the questions preceding and all the questions following the story) are compared, the difference in learning caused by a different placement of the same questions is as great as the difference caused by omitting or including the questions. In fact, it is a trifle greater. In Element IV the mean score of T is one hundred forty-five per cent of the mean score of X, while the greatest difference between the forms with questions and the form without questions is forty-one per cent in Element III. We reach a similar conclusion from the results in Elements VII, VIII and IX—the difference is greater between the forms with questions differently placed than between the forms with and without questions.

Unless Elements II and III are considered more important than all other elements combined, we must come to the somewhat remarkable conclusion that poorly placed questions are worse than no questions at all; and that of the four placements studied, the poorest is that most commonly used—namely, at the end of the chapter (or story).

In general we may say that the placement of questions makes an important difference in the learning of both the material to which the questions directly refer and the material to which they do not directly refer.

3. *Do questions aid the recall of only those facts with which they deal?*

In general it may be said that questions aid only in the recall of the facts with which they deal. (Facts here should be differentiated from generalizations.)

4. *Do questions aid recall by attracting attention to one set of facts at the expense of other facts?*

When all the questions precede the story, recall of the facts connected with these questions is aided, and recall of other facts is not interfered with. When questions are interspersed either at the beginning or end of appropriate paragraphs, the recall of facts connected with these questions is aided, apparently, at the expense of the recall of other facts. Similarly, when all the questions are at the end of the story, there is a gain in the recall of facts connected with the questions and a proportional loss in the recall of other facts.

In general, then, questions do aid recall by attracting attention to one set of facts at the expense of other facts, unless the questions are placed at the beginning of the story.

5. *Do questions dealing with generalizations have any effect upon ability to draw general conclusions from other facts in the story?*

The forms with the questions preceding the story and with the questions interspersed show a gain in generalizations connected with the fact questions in the story corresponding to (although less than) the gain in generalizations connected with the generalization questions. There was little or no change in the non-experimental generalizations. So we may say that preview and interspersed questions dealing with generalizations have no effect upon ability to draw general conclusions from other facts in the story, unless those other facts are emphasized by factual questions, in which case the ability to draw general conclusions from them is increased.

When all the questions follow the story there is a loss in the ability to draw general conclusions from the facts not specifically covered by the generalization questions in the story. This loss is proportional to the gain made in generalizations directly called for by the review questions. To make the contrast sharper, we may say that when facts and generalizations connected with the questions in the story are given equal weight with all the other facts and generalizations, preview questions result in a net gain, whereas review questions result in no gain, improvement in the recall of one set of facts being almost exactly offset by loss in the recall of other facts, and the improvement of one set of generalizations being almost exactly offset by loss in other generalizations.

6. *Do questions which tend to increase generalizing about a story affect favorably, unfavorably, or not at all the recall of factual material?*

So far as the placement of questions goes, the answer to this question is clear: The more an arrangement favors the recall of facts the more it also favors the making of generalizations.

In regard to the difference in the effect of fact questions and generalization questions, the answer is less clear. But this is apparently due to the combined effect of factual and generalization questions acting in opposite—or at least different—directions. No very definite conclusions can be drawn from the present study as to the difference in the effects of these two types of questions.

7. *Should questions be grouped at the end of the chapter, at the beginning of the chapter, or scattered throughout the chapter?*

So far as the results of the present experiment may be generalized, it seems very clear that the best placement of questions is at the beginning of the chapter or story. Such a placement is reliably superior to the other placements in nearly all factors, both experimental and non-experimental, and is reliably inferior in none. It results in a decided gain in the recall of facts and in the making of generalizations directly related to the questions in the story. It results in no appreciable loss in the recall of facts not connected with the questions in the story. And it results in a spread of generalizing ability to facts not covered by questions calling for generalizations. In brief, not only is the recall of specific facts aided, but general comprehension of the significance of the facts is very markedly increased.

Interspersing the questions at the beginning and the end of appropriate paragraphs results in a gain in general comprehension, and in the recall of the facts covered by the questions. But the latter is offset by an equal loss in the recall of facts not covered by the questions. There is no loss in generalizations not covered by the questions.

The results in regard to review questions (all questions grouped at the end of the story) are particularly interesting not only because it is the form most frequently employed by textbook writers, but also because the results strongly hint at the relative efficacy of preview and review in general. The form with review questions was superior to the other forms in no respect whatever. In total score it is the only one of the four (experimental) forms containing questions which falls below the (control) form containing no questions. Only in direct answers to the questions is the form containing review questions superior to the form containing no questions, and even in this respect preview questions resulted in more than twice as much improvement as that effected by the review questions.

8. *Is the effect of the type and placement of questions the same upon children of different grades and ages, boys and girls, the bright and the dull?*

The analysis of the data necessary for an adequate answer to this question is not made in the present study. It may be said, however, that Form T (preview questions) shows the most consistent superiority for both sexes and in all three grades, while Form X shows the most consistent inferiority. In all three grades the differences between the forms is much greater for the girls than for the boys. In particular (as shown in Fig. 1), although boys and girls are both favorably affected by preview questions, the girls are affected (by mind set?)

about twice as much, on an average, as the boys. On the other hand, boys seem to benefit by review questions somewhat more than girls. The older children of both sexes are somewhat more benefited by interspersed questions following appropriate paragraphs than are the younger children.

Referring to those children who are under age for their grade as "bright" and to those who are over age for their grade as "dull," we may say that questions affect the bright children far more favorably than they affect the dull children, but that for both bright and dull children preview questions are superior to review questions.

Summary

1. The inclusion of questions in the historical account of "How the Florentine People Struggled for Democracy" results in an important and reliable difference in the recall and understanding of the story.

2. An equally reliable and important difference results from different placements of the questions.

3. The best placement is the grouping of all questions at the beginning of the story.

4. This placement results in a significant gain in the recall and understanding of the parts of the story covered by the questions, and in no corresponding loss in the recall and understanding of other parts of the story.

5. Most of the gains resulting from this placement (and from the interspersed placements, as well) are traceable to mind set, and are in contrast to the gains resulting from the placement of the questions at the end of the story as a review exercise.

6. The placement of questions at the beginning or at the end of appropriate paragraphs results in a loss in the recall of facts not covered by the questions proportional to the gain in the recall of facts covered by the questions.

7. These two (interspersed) arrangements result in a marked gain in understanding (as distinguished from the recall of facts) of the parts of the story covered by the questions and in no corresponding loss in understanding of the parts of the story not covered by the questions.

8. The worst placement is the grouping of all questions at the end of the story.

9. This placement results in some gain in the recall and understanding of the parts of the story covered by the questions, and in a

proportional loss in both the recall and understanding of all other parts of the story.

10. Contrasting the two preview placements of questions (preceding the story and preceding the appropriate paragraphs) with the two review placements of questions (following the appropriate paragraphs and following the story) the conclusion is that preview is greatly superior to review.

11. The girls appear to be much more affected by mind set (preview questions) than are boys.

12. Boys appear to be somewhat more benefited by review questions than are girls.

13. Children with high EQ's are more affected by questions and by differences in the placement of questions than are children with low EQ's.

14. Preview questions calling for generalizations result in an improvement in generalizing which spreads to facts not covered by such questions.

THE INFLUENCE OF AGE OF PUBESCENCE UPON THE PHYSICAL AND MENTAL STATUS OF NORMAL SCHOOL STUDENTS*

MORRIS S. VITELES

University of Pennsylvania

Differences between pre-pubescents and post-pubescents of the same age have been noted in a number of studies. In 1907, Crampton[3] published observations to the effect that the post-pubescent is superior in height, weight and strength to pre-pubescents of the same age. In addition, he found the scholarship ratings of the post-pubescents to be distinctly superior to those of the pre-pubescents. In a later study, Baldwin[2] concludes that "physiological age is directly correlated with stages of mental maturity—results showing that at each chronological age the physiologically accelerated boys and girls have a higher mental age than those of the average or below the average physiological age." In another study he finds "the mean mental age of a group of physiologically accelerated children to be higher than that of a group of physiologically retarded children." In a still later study of more than one hundred boys, Lutz[5] finds marked differences in age of pubescence among dull, normal and superior boys as determined by the Terman IQ, postpubescence being reached in the superior group at about thirteen years of age, in the average group at the age of fifteen, and in the inferior at sixteen years of age.

Figures obtained by Terman *et al.*[6] in a study of gifted childen indicate that gifted boys and girls tend to mature earlier than average. Significant relationships are likewise found between the degree of maturation and height and weight respectively.

The findings discussed above are somewhat in contradiction to those following from a study of over eight-hundred girls in the Chicago Schools by Abernethy,[1] who concludes that there is "little or no relationship between mental age and the generally accepted criteria of physiological and anatomical development." This holds not only for the entire group but when the extremes in age of pubescence are compared. In the case of height and weight the author finds no evidence that these are ultimately influenced by age of maturing, except in instances of extreme retardation in maturing. On the other hand,

* The raw data forming the basis of this paper were assembled by Miriam C. Simon, Genevieve McDermott, and E. G. Kynett, graduate students in psychology.

in a study of grade location, class standing, physical health, social development and adjustment to the demands of the school, the extremely physiologically accelerated girls are shown to be superior to the physiologically retarded, suggesting that "physiological maturing has bearing upon important intellectual, emotional and volitional traits not measurable by intelligence tests."

Gesell[4] likewise reaches the conclusion that pubertal precocity does not appear to result in acceleration of mental growth, the changes which take place concerning personality as contrasted with intellectual factors. In general, however, early pubescence is found by this investigator to have a more favorable influence in general mental development than delayed pubescence.

Purpose of the Present Study

All of these studies are limited to an examination of the relationship between diverse variables and the age of maturity of boys and girls who have very recently attained or who are about to attain maturity. In all the groups studied, including Abernethy, there are

Table I.—Number of Years Elapsing since Attainment of Maturity

0 to 11 months	1
1 year to 1 year., 11 months	6
2 years to 2 years, 11 months	8
3 years to 3 years, 11 months	38
4 years to 4 years, 11 months	82
5 years to 5 years, 11 months	60
6 years to 6 years, 11 months	24
7 years to 7 years, 11 months	9
8 years to 8 years, 11 months	2
9 years to 9 years, 11 months	3
10 years to 10 years, 11 months	1
11 years to 11 years, 11 months	1
12 years to 12 years, 11 months	1
	236

individuals who have not yet matured sexually. The subjects of the present study are two hundred thirty-six first-year students, ranging in chronological age from fifteen years, eleven months to twenty-four years, in a Normal School for women situated in a large Eastern city. The investigation represents an attempt to establish in a preliminary way the extent to which early maturity confers a benefit in the way of superior mental status, height, weight, etc., extending into adult life.

In contrast with the subjects of the investigations cited above, all in the present investigation have attained maturity. As a matter of fact, in the case of all except seven girls, over two years have elapsed since pubescence, and there is a median lapse of four years. (See Table I.)

Source of Data

For these subjects present height, weight, academic standing and intelligence levels have been compared with age of pubescence. Height and weight were taken from the records of the medical examination given by the Physical Education Department. Academic achievement has been measured by the average of academic grades for the first half year of instruction in the Normal School. The raw score on the Brown University Psychological Examination was taken as the index of general intelligence.

Age of first menstruation as furnished by each girl was taken as the age of pubescence. The physical directress of the school, a physician, asked each girl to furnish this information in the course of a medical examination. In cases of doubt on the part of the girl as to the exact date she was instructed to consult her mother.

A note concerning this method seems in place. There are, undoubtedly, possibilities for error in such a procedure. Probability of exact recall after a lapse of from two to twelve years may, in general, be small. However, the event to be recalled in this case is so outstanding in the life of the girl that the probability for correct recall

Table II.—Ages of Maturing of Two Hundred Thirty-six Normal School Students (Viteles) and of Four Hundred Eighty-seven Girls in the University of Chicago High Schools (Abernethy)

Age	(Abernethy) (high school)		(Viteles) (normal school)	
	Number of cases	Per cent	Number of cases	Per cent
10	5	1.0	0	0
11	37	7.7	14	6.0
12	112	23.0	55	23.3
13	172	35.3	84	35.6
14	123	25.2	59	25.0
15	29	6.0	13	5.5
16	8	1.6	6	2.5
17	1	.2	5	2.1
Total.....	487	....	236	

is doubtless increased. In addition, consultation with the mother in cases of doubt can be expected to increase the accuracy of the information supplied by the girls. Moreover, women were selected as subjects because of the objectivity of the symptoms accompanying maturation, which can likewise be expected to contribute to the accuracy of the dates given. As a matter of fact, it was found that the figures given by the girls distributed themselves into a normal curve coinciding closely with the distribution of ages of pubescence in Abernethy's study (see Table II). Age of pubescence was found to vary from nine, to seventeen years six months, the mean being 13.56 ± 1.25 and the median 13.5 years, figures which also conform to those obtained by Abernethy.

Results of Statistical Treatment of Data

The data obtained in this manner have been treated in a number of ways with a view of examining the relationships among the variables described above in the case of these post-pubescent girls.

1. A comparison has been made of the *average* height, weight, intelligence test score and chronological age of the girls in the lowest, lower, middle, higher and highest quintile from the point of view of age of pubescence (See Table III).* Briefly summarized, the findings of this analysis are as follows:

(A) There are no consistent differences between higher and lower quintiles in *height, intelligence test score* and *academic grade.* In fact, the girls in Quintile V, *i.e.*, the twenty per cent who mature latest, average .6 of an inch taller in *height* $\left(\frac{D}{\text{SD difference}} = 1.2\right)$ than those in in Quintile I, *i.e.*, the twenty per cent who mature earliest, and make an average of .4 of a point more on the intelligence test $\left(\frac{D}{\text{SD difference}} = 0.20\right)$. The average *academic grade* of the lowest quintile is 81.9; of the highest 81.3, a difference of only .6 of a point in favor of those maturing earliest $\left(\frac{D}{\text{SD difference}} = .72\right)$.

(B) In *weight* there is a persistent drop in average from Quintile II, ranging from 124.1 lbs, for Quintile II; 121 lbs, for Quintile III; 116.6 lbs. for Quintile IV; to 113.8 lbs, for the girls in Quintile V.

* Quintile I includes the twenty per cent of girls maturing earliest, *i.e.*, from year to 9.0 to 12.5; Quintile V includes the twenty per cent maturing latest, *i.e.*, from 14.3 to 17.6.

TABLE III.—COMPARISON OF PUBESCENT AGE (QUINTILE DISTRIBUTION) AND AVERAGE HEIGHT, WEIGHT, INTELLIGENCE AND ACHIEVEMENT SCORE

Quintiles	Number of cases	Age of pubescence, range in years	Average age of pubescence in years	Average chronological age in years	Average height in inches	Average weight in pounds	Average raw score Brown University PE	Average grades first term normal school
			σ	σ	σ	σ	σ	σ
I	48	9.0 to 12.5	11.89 ± .60	18.3 ± 1.05	62.2 ± 2.35	119.6 ± 14.5	47.4 ± 11.10	81.9 ± 4.10
II	48	12.6 to 13.0	12.90 ± .22	18.1 ± 1.07	62.9 ± 2.22	124.1 ± 18.7	46.2 ± 7.68	82.2 ± 3.80
III	45	13.1 to 13.8	13.43 ± .13	18.1 ± 1.51	62.4 ± 2.66	121.0 ± 20.5	44.5 ± 9.35	81.6 ± 4.05
III	44	13.9 to 14.2	13.95 ± .11	18.2 ± .74	62.0 ± 2.64	116.6 ± 12.8	45.9 ± 7.00	82.9 ± 4.33
V	51	14.3 to 17.6	15.13 ± .86	18.7 ± 1.23	62.8 ± 2.42	113.8 ± 10.2	47.8 ± 8.60	81.3 ± 4.11
Entire group	236	9.0 to 17.6	13.56 ± 1.25	18.3 ± 1.20	62.4 ± 2.42	118.5 ± 15.8	46.2 ± 8.88	81.6 ± 4.08

The difference of 10.3 lbs. between Quintiles II and V is completely reliable $\left(\frac{D}{\text{SD difference}} = 3.4\right)$. On the other hand, the average weight of the twenty per cent who mature earliest is 119.6, only slightly above the average for the entire group. However, these figures on *weight* may be said to confirm somewhat the opinion expressed by Abernethy[1] on the possible influence of extremely late maturing on physical growth.

(C) There is no significant trend of differences in average *chronological age* among the quintiles. The average chronological age of the twenty per cent of the girls who mature earliest (Quintile I) is 18.3: the girls who mature latest (Quintile V) average 18.7 years in age. $\left(\frac{D}{\text{SD difference}} = 1.7.\right)$ The middle twenty per cent average 18.1 years in chronological age. The significance of these figures lies in the indication that those who mature early do not enter normal school at an earlier chronological age that those who mature at a later age.

II. With the thought that extremes in physiological retardation and acceleration might show related co-variability in other traits, a comparison was made of sixteen students maturing from nine to eleven years, eleven months of age with fourteen students maturing between fifteen years and seventeen years, six months of age, all the cases at the two extremes, constituting 12.6 per cent of the entire group. The results of this comparison are presented in Table IV, showing the averages of the two groups on intelligence test score, achievement score, etc., the SD, SD difference and the $\frac{D}{\text{SD difference}}$, indicating the reliability of the difference between the two groups on each trait. The analysis shows no significant differences on any trait between groups so extreme from the point of view of physiological acceleration and retardation.

III. The relationship between pubescent age and the remaining variables has been further studied by the method of correlation. (See Table V.) In no case is the coefficient high enough to demonstrate the existence of co-variability between age of pubescence and any of the other factors considered in this investigation. This holds true when chronological age is made constant by the method of partial correlation as well as for the zero order coefficients.

TABLE IV.—COMPARISON OF PHYSIOLOGICALLY ACCELERATED AND RETARDED NORMAL SCHOOL STUDENTS

Age of pubescence, range in years	Number of cases	Average age of pubescence in years, σ	Average chronological age in years, σ	Average height in inches, σ	Average weight in pounds, σ	Average score Brown University intelligence examination, σ	Average grade, first term normal school, σ
9.0 to 11.11 (accelerated).....	16	10.9 ± .91	18.4±1.7	62.2±2.02	119.0±15.9	47.8±13.4	82.2±4.4
15.1 to 17.6 (retarded).....	14	16.0 ± .64	18.6±1.35	62.8±2.40	114.6± 9.2	47.4± 5.9	81.5±3.7
Entire group.....	236	13.56±1.25	18.3±1.25	62.4±2.42	118.5±15.8	46.2± 8.88	81.6±4.08
Difference (accelerated and retarded).....	...	−5.1	−.20	−.60	+4.4	+.40	+.70
SD difference.....	...	1.74	.55	.82	4.6	3.7	1.48
$\frac{\text{Difference}}{\text{SD difference}}$	...	2.9	.36	.73	.95	.13	.47

Summary of Findings and Conclusions

No significant relationship between age of pubescence and height, weight, intelligence and academic achievement exists in the case of two hundred thirty-six post-pubescence girls in the first year of the normal school.

In the light of differences between younger pre-pubescents and post-pubescents reported in earlier studies (2, 3, 5) the findings of the present investigation have certain important implications.

I. The findings indicate that such advantages in height, weight, intelligence, etc., as *may* be associated with early pubescence are merely temporary in character. The results of earlier studies have led to the conclusion that pubescence is accompanied by a spurt in physical growth and mental development. The findings of the present investigation suggest that those who mature late profit from this spurt equally as much as those who mature early, the effect of the spurt in the two groups being equalized as adult life is approached.

Table V.—Correlation of Age of Pubescence with Chronological Age, Height, Weight, Etc.

1. Age of Pubescence
2. Height
3. Weight
4. Chronological age
5. Brown University intelligence examination score
6. Academic achievement

$r12 = +0.09 \pm .043$	
$r13 = -0.14 \pm .042$	$r12.4 = +0.08$
$r14 = +0.11 \pm .043$	$r13.4 = -0.13$
$r15 = +0.01 \pm .043$	$r15.4 = +0.02$
$r16 = -0.04 \pm .043$	$r16.4 = -0.02$

This conclusion can only be formulated tentatively. Its validity must be checked on other less highly selected groups, students admitted to normal school being highly selected, both from point of view of intellectual status and physical appearance, in which height and weight are factors. The genetic approach involving a follow up of pre-pubescence through pubescence into maturity is the most satisfactory method of investigating further the effect of pubescence upon physical and mental growth.

II. These findings have significant implications in individual diagnosis at the pubescent level. To determine mental age and academic status at this level is not enough. Physiological age must be

established with a view of determining whether the subject is still to profit from the spurt in mental development which seems to accompany pubescence, or whether this possible avenue of adjustment is closed. Instead of a source of additional pride to the bright, the stimulating effect of pubescence, regardless of the extent to which its appearance is delayed, becomes a source of hope for the older dull pre-pubescent.

III. The present investigation is incomplete in its failure to include a study of differences in emotional traits and social adjustments of the physiologically retarded and accelerated, the further investigation of which would appear from the studies of Abernethy[1] and Gesell[4] to be a fertile source of important data.

Bibliography

1. Abernethy, E. M.: Correlations in Physical and Mental Growth, *Journal Educational Psychology*, 1925, Vol. XVI, pp. 458–466, 539–546.
2. Baldwin, Bird T.: "The Physical Growth of Children from Birth to Maturity." *University of Iowa Studies in Child Welfare*, 1921, p. 1.
3. Crampton, C. Ward: The Influence of Physiological Age upon Scholarship. *Psychological Clinic*, 1907, Vol. I, pp. 115–120.
4. Gesell, A.: The Influence of Puberty Praecox upon Mental Growth. *Genetic Psychological Monograph*, 1926, Vol. I, No. 6, pp. 511–539.
5. Lutz, Wm. F.: The Relation of Mental to Physical Growth. *Psychological Clinic*, 1924, Vol. XV, pp. 125–129.
6. Terman, Lewis M. et al.: "Genetic Studies of Genius." *Stanford University Press*, Vol. I, 1925, 648 pages.

ON FORMULA FOR ESTIMATING THE RELIABILITY OF TEST BATTERIES

HARL R. DOUGLASS

University of Oregon

AND

FREDERICK W. COZENS

University of California at Los Angeles

The usefulness of the Spearman-Brown prophecy formula, $r_{af,Af} = \frac{ar_{11}}{1 + (a-1)r_{11}}$, in estimating the reliability of a composite of a number of phases or a number of judgments of the same function has been shown.[1,2,3] It is significant to note that the Spearman-Brown prophecy formula is derived from Spearman's original formula for the correlation between sums or averages[4] by assuming a equal to b and $r_{pQ} = r_{pq} = r_{PQ}$ and equal to the average of the reliability coefficients of the separate parts of the test battery. The r_{pQ} refers to the intercorrelation of all tests plus the reliability coefficients of the items; r_{pq} to the intercorrelations only; and r_{PQ} to the intercorrelations of Form II of the tests. This equality may be safely assumed in those instances in which tests 1 to a and I to b are all *comparable and similar forms of the same test.*

It is often desirable to estimate reliability coefficients of batteries of tests when the intercorrelations of the single tests or elements of a battery may not be safely regarded as equivalent in magnitude to

---

[1] Holzinger, Karl J.: Note on the Use of Spearman's Prophecy Formula for Reliability. *Journal of Educational Psychology*, Vol. XIV, May, 1923, pp. 302–305.

[2] Holzinger, Karl J.: Further Experiments in the Application of Spearman's Prophecy Formula. *Journal of Educational Psychology*, Vol. XVI, May, 1925, pp. 289–299.

[3] Kelley, T. L.: Applications of the Spearman-Brown Formula for the Measurement of Reliability. *Journal of Educational Psychology*, Vol. XVI, May, 1925, pp. 300–303.

[4] See Kelley, T. L.: "Statistical Method." Formula 147, p. 197.

$$r_{(1+2+\cdots a)(I+II+\cdots b)} = \frac{\overset{ab}{\underset{1}{S}} r_{pQ}}{\sqrt{a + \overset{a^2-a}{\underset{1}{S}} r_{pq}}\sqrt{b + \overset{b^2-b}{\underset{1}{S}} r_{PQ}}}$$

the average of the reliability coefficients. In such an instance it would seem logical to use a modification of the Spearman formula for the correlation between the sums or averages of weighted scores. Kelley[1] gives this formula as

$$r\left(\sum_{1}^{a} w_p X_p\right)\left(\sum_{1}^{b} w_P X_P\right) = \frac{\sum_{1}^{ab} w_p \sigma_p w_Q \sigma_Q r_{pQ}}{\sqrt{\sum_{1}^{a} w_p^2 \sigma_p^2 + \sum_{1}^{a^2-a} w_p \sigma_p w_q \sigma_q r_{pq}}\sqrt{\sum_{1}^{b} w_P^2 \sigma_P^2 + \sum_{1}^{b^2-b} w_P \sigma_P w_Q \sigma_Q r_{PQ}}} \quad (1)$$

A special and simpler statement of this formula may be made for use in situations in which certain special but frequently occurring conditions obtain. In order to make the simplification of the formula and the solution of a problem more clear, a specific instance will be cited. In a study of general athletic ability in college men,[2] it was necessary to determine the reliability of an element of general athletic ability containing four sub-tests. Data for the calculations are found in Table I.

TABLE I.—INTERCORRELATIONS FOR THE ELEMENT, *Speed of Legs*

Sub-test	Reliability coefficient	Dodging	Medicine ball run	120 yard low hurdles	100 yard dash
Dodging.....	.850	....	.602	.565	.500
Medicine ball run....	.643	.602	....	.550	.546
120 yard low hurdles.	.945	.565	.550	....	.633
100 yard dash.....	.968	.500	.546	.633	

$r_{af.Af}$ = .958. Average r_{11} = .8515, which is clearly much higher than the average of the intercorrelations.

The significance of the notations of the formula in terms of the problem is as follows:

$a = b = 4 =$ IV = number of tests to be summed or averaged.

$w_p =$ the nominal weight to be assigned each test in the group in obtaining the sum or average (Form I), p referring in turn to each test from 1 to 4.

[1] Kelley, T. L.: *Op. cit.*, formula 148, p. 198.

[2] Cozens, F. W.: "The Measurement of General Athletic Ability in College Men." University of Oregon Publications. Physical Education Series, Vol. I, No. 3, April, 1929.

w_Q = as above for Form II, Q referring in turn to each test from I to IV.

σ_p = the standard deviation of each test in turn in the group from 1 to 4 (Form I).

σ_Q = the standard deviation of each test in turn in the group from I to IV (Form II).

r_{pQ} = the correlation between each test in the group (Form I) and every test in the group (Form II), *i.e.*, all the intercorrelations plus the reliability coefficients.

w_q = the weight of every test in one form in turn except that test being considered at the particular time in w_p, *i.e.*, $\underset{1}{\overset{a}{\mathrm{S}}} w_p w_q$ refers to the sum of the products of the weight of each test in one form multiplied in turn by the weight of every other test in that form.

w_P has a similar significance for the tests of Form II.

σ_q similarly refers to the standard deviation of every test other than that being considered at the particular time in σ_p, *i.e.*, $\underset{1}{\overset{a}{\mathrm{S}}} \sigma_p \sigma_q$ = the sum of the products or the standard deviations of every test multiplied by the standard deviations of every other test in that form.

σ_P has a similar significance for the tests of Form II.

$\underset{1}{\overset{a}{\mathrm{S}}} r_{pq}$ refers to the intercorrelations of the tests in Form I not including the reliability coefficients.

$\underset{1}{\overset{a}{\mathrm{S}}} r_{PQ}$ has the same significance as r_{pq} except that it refers to the tests of Form II.

Since the data of the study mentioned are expressed in T-Scale units, the standard deviations of all tests will be equal theoretically to ten. This is actually true except for small and insignificant differences due to chance. Thus $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_I = \sigma_{II} = \sigma_{III} = \sigma_{IV}$. The tests in the element were equally weighted. Hence the factor $w_p w_Q \sigma_p \sigma_q$ ($= w_p{}^2 \sigma_p{}^2 = w_p \sigma_p w_q \sigma_q = w_P{}^2 \sigma_P{}^2 = w_P w_Q \sigma_P \sigma_Q$) will cancel from the formula leaving

$$r\left(\underset{1}{\overset{a}{\mathrm{S}}} X_p\right)\left(\underset{1}{\overset{b}{\mathrm{S}}} X_P\right) = \frac{\underset{1}{\overset{ab}{\mathrm{S}}} r_p Q}{\sqrt{a + \underset{1}{\overset{a^2-a}{\mathrm{S}}} r_{pq}}\sqrt{b + \underset{1}{\overset{b^2-b}{\mathrm{S}}} r_{PQ}}}$$

By the nature of the data a equals b. Since the reliability coefficients of the sub-tests were determined from different groups, the intercorrelations r_{PQ} are not available but, as they refer to tests exactly similar and comparable to those referred to by r_{pq}, it may be assumed that r_{PQ} is a very close approximation to and theoretically identical with r_{pq}, *i.e.*, $r_{12} = r_{\text{I II}}$, $r_{13} = r_{\text{I III}}$, . . . $r_{34} = r_{\text{III IV}}$. Substituting these values,

$$r\left(\sum_{1}^{a} X_p\right)\left(\sum_{1}^{a} X_P\right) = \frac{\sum_{1}^{a^2} r_{pQ}}{a + \sum_{1}^{a^2-a} r_{pq}}$$

Of those correlations to be summed as r_{pQ}, the reliability coefficients $r_{1\,\text{I}}$, $r_{2\,\text{II}}$, etc., are available, but the intercorrelations $r_{1\,\text{II}}$, $r_{1\,\text{III}}$, etc., are lacking. However, theoretically equal to and practically very closely approximate to them are the intercorrelations r_{pq}, *i.e.*, r_{12}, r_{13}, etc. The formula as actually used then is,

$$r\left(\sum_{1}^{a^1} X_p\right)\left(\sum_{1}^{a} X_P\right) = \frac{\sum_{1}^{a} r_{pP} + \sum_{1}^{a^2-a} r_{pq}}{a + \sum_{1}^{a^2-a} r_{pq}} \qquad (2)$$

Substituting in this formula the appropriate values, the estimated reliability coefficient of the element *Speed of Legs* equals .944 as against an estimated coefficient of .958 by means of the Spearman-Brown prophecy formula.

To show further how the Spearman-Brown prophecy formula may overstate the reliability coefficient, we cite again from the study on general athletic ability in college men:

TABLE II.—COMPARISON OF ESTIMATES OF RELIABILITY[1]

Elements of general athletic ability	Number of sub-tests in element	Average reliability coefficient of sub-tests	Average of intercorrelations	Range of intercorrelations	Reliability coefficient of element by S-B formula	Reliability coefficient of element by formula (2)
Arm and shoulder-girdle strength....	6	.894	.483	.368–.585	.981	.969
Arm and shoulder-girdle coordination.....	5	.913	.627	.625–.728	.981	.976
Hand-eye, foot-eye, arm-eye coordination.....	8	.493	.289	.127–.432	.886	.831
Jumping and leg strength.....	8	.929	.337	.156–.615	.990	.984
Endurance.....	3	.837	.460	.247–.804	.939	.909
Body coordination, agility and control	7	.836	.338	.271–.405	.973	.947
Speed of legs.....	4	.851	.566	.500–.633	.958	.944

[1] These and other similar data are taken from Dr. Cozens' dissertation, *op. cit.*

USE OF THE FORMULA WITH WEIGHTED BATTERY SCORES

Suppose now that we wish to formulate a short battery of tests which will measure best all the elements of general athletic ability. In order that this battery may correlate as high as possible with the criterion, the sub-tests were weighted. Weights were derived by Kelley's method of successive approximations.[1] The weights of the sub-tests are not now equal as they were in the previous instance but the standard deviations are by the nature of the data (raw scores being reduced to T-Scores. Also we may assume that $w_p = w_P$ and $w_q = w_Q$, that is, $w_p w_q = w_P w_Q$. Hence when weighted sub-tests are used, formula (1) reduces to

$$r\left(\sum_{1}^{a} w_p X_p\right)\left(\sum_{1}^{a} w_P X_P\right) = \frac{\sum_{1}^{a^2} w_p w_Q r_{pQ}}{\sum_{1}^{a} w_p^2 + \sum_{1}^{a^2-a} w_p w_q r_{pq}}$$

[1] Kelley, T. L.: *Op. cit.*, sec. 84, p. 302.

In the form of formula (2) this equation becomes

$$r\left(\sum_{1}^{a} w_p X_p\right)\left(\sum_{1}^{a} w_P X_P\right) = \frac{\sum_{1}^{a} w_p^2 r_{pP} + \sum_{1}^{a^2-a} w_p w_q r_{pq}}{\sum_{1}^{a} w_p^2 + \sum_{1}^{a^2-a} w_p w_q r_{pq}} \qquad (3)$$

In the situation at hand the sub-tests with reliability coefficients and weights are as follows:

TABLE III.—RELIABILITY COEFFICIENTS AND WEIGHTS OF SUB-TEST

Sub-test	Reliability coefficient	Weight	Weight squared	$w_p^2 r_{pP}$
Dip (on parallel bars)	.926	.6	.36	.33336
Baseball throw for distance	.912	1.5	2.25	2.05200
Football punt for distance	.816	1.0	1.00	.81600
Standing broad jump	.965	1.0	1.00	.96500
Quarter-mile run	.917	1.2	1.44	1.32048
Long dive (dive for distance)	.930	.7	.49	.45570
Dodging	.850	1.0	1.00	.85000
Sums	.....	....	7.54	6.79254
Average relative coefficient	.902			

$$\sum_{1}^{a^2-a} w_p w_q r_{pq} = 15.71122.$$

Substituting in formula (3),

$$r_{11\ (\text{battery})} = \frac{6.79254 + 15.71122}{7.54 + 15.71122} = .968.$$

By the Spearman-Brown formula the reliability coefficient of the battery would have been estimated as .985.

It may be interesting to point out the difference in estimation between the two methods in a number of cases and with varying numbers of sub-tests in a battery. Table IV illustrates the spuriously high reliability obtained by the Spearman-Brown formula.

TABLE IV.—ESTIMATION OF WEIGHTED BATTERY RELIABILITY COEFFICIENTS

Battery number	Number of tests in battery	Average reliability coefficient of sub-tests	Battery reliability coefficient by S-B formula	Battery reliability coefficients by formula (3)	Differences
1	7	.902	.985	.968	.017
2	7	.898	.984	.969	.015
3	6	.916	.985	.965	.020
4	6	.920	.986	.971	.015
5	5	.931	.985	.970	.015
6	5	.916	.982	.958	.027
7	4	.911	.976	.960	.016
8	4	.945	.986	.975	.011
9	4	.935	.983	.968	.015
10	4	.931	.982	.965	.017
11	4	.904	.974	.944	.030
12	5	.911	.981	.961	.020

It is obvious that the discrepancy between coefficients estimated by means of the Spearman-Brown formula and those estimated by means of the Spearman formula for the correlation between sums or averages is a function of the number of tests or sub-tests and the difference between the average of the reliability coefficients and the average of the intercorrelations. Let us consider hypothetical cases in which such differences are greater than those obtaining in the data given above.

TABLE V.—RELIABILITY COEFFICIENTS ESTIMATED BY MEANS OF (1) S-B PROPHESY FORMULA AND (2) SPEARMAN FORMULA FOR SUMS AND AVERAGES

Number of tests in battery	Average reliability coefficient of sub-tests	Average intercorrelation of sub-tests	Battery reliability coefficient by S-B formula	Battery reliability coefficient by formula (2)	Differences
2	.800	.600	.889	.875	.014
5	.800	.600	.953	.942	.011
10	.800	.600	.976	.969	.007
2	.800	.400	.889	.858	.031
5	.800	.400	.953	.924	.029
10	.800	.400	.976	.957	.019
2	.800	.200	.889	.833	.056
5	.800	.200	.953	.889	.064
10	.800	.200	.976	.929	.047
2	.500	.100	.666	.545	.121
5	.500	.100	.833	.643	.190
10	.500	.100	.909	.732	.177

It will be noted from Table V that the difference between the estimated correlation obtained by use of the S-B formula and that given by formula (2) increases with the difference between the average reliability coefficient of the sub-tests and the average intercorrelation of the sub-tests; that it decreases with the increase of the number of sub-tests; and that it increases with the decrease of the average reliability coefficient, other things remaining constant. An examination of the two formulas will show that these facts necessarily follow the nature of the formulas, *e.g.*:

$$r(\text{estimated by means of S-B formula}) = \frac{ar_{pP}}{1 + (a-1)r_{pP}}$$

$$r(\text{estimated by means of formula (2)}) = \frac{r_{pP} + (a-1)r_{pq}}{1 + (a-1)r_{pq}}$$

It will readily be seen that as r_{pq}, the average intercorrelation, approaches r_{pP}, the average reliability coefficient, the two formulas become the same. As a increases, r estimated by formula (2) approaches unity, and since r estimated by the S-B formula always lies between that estimated by (2) and unity, the difference between the values estimated for r by the two formulas is certain to decrease with the increase in a.

Examination of the two formulas should also reveal that the more greatly r_{pP}, the average reliability coefficient, differs from 1, the greater will be the influence of the discrepancy between the average reliability coefficient and the average intercorrelation coefficient, and consequently the greater difference between values for r estimated by the formulas in question.

Conclusions

1. In computing the reliability of test batteries, care should be observed in the use of the Spearman-Brown prophecy formula. Unless the sub-tests are quite similar in "measuring" capacity as demonstrated by high intercorrelations, the Spearman-Brown formula is not applicable.

2. The formula for use in computing the reliability coefficients of batteries, the sub-tests of which incompletely measure phases of the same general ability, is Spearman's formula for the correlation between sums or averages. This is given by Kelley as (1) in the body of the paper. A useful adaptation of this formula with unweighted scores and with equal standard deviations is given as (2).

3. With weighted battery sub-tests, another adaptation of the Spearman formula is possible. This is given as formula (3). Standard deviations are again assumed to be equal.

4. Table II indicates that with low intercorrelations among the sub-tests, a spuriously high reliability coefficient of at least as much as .190 may occur when reliability coefficients are computed by the Spearman-Brown formula. With fairly high intercorrelations, reliability coefficients are not seriously affected.

The difference between the estimates of r yielded by the two formulas, (1) increases with the difference between the average reliability coefficient and the average intercorrelation of the sub-tests, (2) decreases with the increase of the number of sub-tests, and (3) decreases with the increase of the average reliability coefficient of the sub-tests.

Just how great the difference may be is not known and does not lie within the scope of this paper. This phase of the problem, together with the resulting effect upon probable errors of estimate of such differences, will be treated in a forthcoming paper.

A SPECIAL CASE OF THE MULTIPLE CORRELATION COEFFICIENT

A. K. KURTZ

Ohio State University

Often before beginning an investigation in which one is planning to ascertain the interrelations of a number of variables in order to obtain finally a multiple correlation coefficient between the "independent"[1] variables and some outside criterion, it is desirable to obtain a preliminary idea of the expected result in order that the worth of the investigation may be judged more accurately. For example, one may wish to select machine lathe operators by the use of some five or six tests such as: (1) Arithmetical computation, (2) a test of activity as opposed to laziness, (3) speed of various types of reaction, (4) a test of ability to follow the directions and processes indicated on blueprints, and (5) a composite mechanical ability test. Now one might know or be willing to suppose that the intercorrelations of these tests would probably be in the neighborhood of +.20 to +.40, and that the correlations with such a criterion as production measured some time after employment might run a little higher—say, +.30 to +.50. Clearly, if one were to ascertain the limiting values of the possible multiple correlations, one would have a fairly definite idea of what might be expected from such a study before any of the tests had been used (or possibly even before they had been constructed).

In theoretical work, also, we may desire to know what influence n variables, acting with supposed equal effect, would produce upon some criterion or composite variable which includes all or parts of the "independent" variables. In this special case all the correlations with the criterion would equal one value, r_0, and all other intercorrelations would equal another value, r_i, which might or might not equal r_0.

The solution by determinants of the multiple correlations coefficient under such circumstances becomes extremely simple.

Let r_0 = the correlation of any variable with the criterion, X_0;

r_i = any one of the correlations among the n other ("independent") variables;

[1] Throughout this article, the word "independent" enclosed in quotation marks is used to denote variables which are used to predict a certain criterion variable. The word does *not* imply mathematical independence.

$$c = \frac{r_0}{r_i};$$

$n + 1$ = the total number of variables including the criterion variable;

and

$R_{0.123\ldots n}$ = the multiple correlation coefficient $= \sqrt{1 - \frac{\Delta}{\Delta_{00}}}$

$$\text{where } \Delta = \begin{vmatrix} 1 & r_{01} & r_{02} & r_{03} & \ldots & r_{0n} \\ r_{01} & 1 & r_{12} & r_{13} & \ldots & r_{1n} \\ r_{02} & r_{12} & 1 & r_{23} & \ldots & r_{2n} \\ r_{03} & r_{13} & r_{23} & 1 & \ldots & r_{3n} \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ r_{0n} & r_{1n} & r_{2n} & r_{3n} & \ldots & 1 \end{vmatrix}$$

and Δ_{00} = the minor obtained by deleting the 0th row and the 0th column.

$$\text{Then } \frac{\Delta}{\Delta_{00}} = \frac{\begin{vmatrix} 1 & r_0 & r_0 & r_0 & \ldots & r_0 \\ r_0 & 1 & r_i & r_i & \ldots & r_i \\ r_0 & r_i & 1 & r_i & \ldots & r_i \\ r_0 & r_i & r_i & 1 & \ldots & r_i \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ r_0 & r_i & r_i & r_i & \ldots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_i & r_i & \ldots & r_i \\ r_i & 1 & r_i & \ldots & r_i \\ r_i & r_i & 1 & \ldots & r_i \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ r_i & r_i & r_i & \ldots & 1 \end{vmatrix}}.$$

In the numerator, by subtracting c times the nth column from the 0th column and 1 times the nth column from the 1st, 2nd, . . . , $(n - 1)$th columns; and in the denominator, by subtracting the nth column from each of the other columns, we obtain

$$\frac{\Delta}{\Delta_{00}} = \frac{\begin{vmatrix} 1-c^2r_i & 0 & 0 & 0 & \cdots & 0 & cr_i \\ 0 & 1-r_i & 0 & 0 & \cdots & 0 & r_i \\ 0 & 0 & 1-r_i & 0 & \cdots & 0 & r_i \\ 0 & 0 & 0 & 1-r_i & \cdots & 0 & r_i \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1-r_i & r_i \\ cr_i - c & r_i - 1 & r_i - 1 & r_i - 1 & \cdots & r_i - 1 & 1 \end{vmatrix}}{\begin{vmatrix} 1-r_i & 0 & 0 & \cdots & 0 & r_i \\ 0 & 1-r_i & 0 & \cdots & 0 & r_i \\ 0 & 0 & 1-r_i & \cdots & 0 & r_i \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1-r_i & r_i \\ r_i - 1 & r_i - 1 & r_i - 1 & \cdots & r_i - 1 & 1 \end{vmatrix}}.$$

By adding the 1st + 2nd + $\cdots$ + $(n-1)$th rows to the nth row in both numerator and denominator,

$$\frac{\Delta}{\Delta_{00}} = \frac{\begin{vmatrix} 1-c^2r_i & 0 & 0 & 0 & \cdots & 0 & cr_i \\ 0 & 1-r_i & 0 & 0 & \cdots & 0 & r_i \\ 0 & 0 & 1-r_i & 0 & \cdots & 0 & r_i \\ 0 & 0 & 0 & 1-r_i & \cdots & 0 & r_i \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1-r_i & r_i \\ cr_i - c & 0 & 0 & 0 & \cdots & 0 & 1+(n-1)r_i \end{vmatrix}}{\begin{vmatrix} 1-r_i & 0 & 0 & \cdots & 0 & cr_i \\ 0 & 1-r_i & 0 & \cdots & 0 & r_i \\ 0 & 0 & 1-r_i & \cdots & 0 & r_i \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1-r_i & r_i \\ 0 & 0 & 0 & \cdots & 0 & 1+(n-1)r_i \end{vmatrix}}.$$

which, when summed with respect to the 0th column in the numerator and the 1st column in the denominator, becomes

$$\frac{\Delta}{\Delta_{00}} = \frac{\begin{array}{r}(1 - c^2r_i)(1 - r_i)^{n-1}(1 + nr_i - r_i) \\ + (-1)^{n-1}(cr_i - c)(-1)^{n-2}(1 - r_i)^{n-1}(cr_i)\end{array}}{(1 - r_i)^{n-1}(1 + nr_i - r_i)}$$

Factoring out $(1 - r_i)^{n-1}$ and remembering that $c = \frac{r_0}{r_i}$, we get

$$\frac{\Delta}{\Delta_{00}} = \frac{\left(1 - \frac{r_0^2}{r_i}\right)(1 + nr_i - r_i) - \left(r_0 - \frac{r_0}{r_i}\right)(r_0)}{1 + nr_i - r_i}$$

which reduces to

$$\frac{\Delta}{\Delta_{00}} = \frac{1 + nr_i - r_i - nr_0^2}{1 + nr_i - r_i} = 1 - \frac{nr_0^2}{1 + (n - 1)r_i}$$

Finally,

$$R_{0.123\ldots n} = \sqrt{1 - \frac{\Delta}{\Delta_{00}}} = \sqrt{\frac{nr_0^2}{1 + (n-1)r_i}}$$

Although this formula was derived as a special case of the multiple correlation coefficient, it is, as all multiple correlations are, a correlation between one variable and a (best) weighted sum of a number of "independent" variables. Now it is interesting to note that the right member of this formula is identical with that of the correlation between a criterion and the sum or average of a number of *equally* weighted test scores (as derived by Spearman[1] and quoted by Kelley[2] and Holzinger[3]). This identity of formulæ can only mean that the (β) regression weights in a multiple correlation between such tests are all equal and, without loss of generality, can all be assumed equal to 1.

Let us now consider the problem mentioned in the first paragraph of this article. Substituting in the formula values of r_i from .20 to .40, values of r_0 from .30 to .50, and letting $n = 5$, we obtain values

---

[1] Spearman, C.: Correlations of Sums or Differences. *British Journal of Psychology*, 1913, Vol. V, pp. 417–426. Formula (13).

[2] Kelley, Truman L.: "Statistical Method." Formula (151).

[3] Holzinger, Karl J.: "Statistical Methods for Students in Education." Formula (51). Dr. Holzinger also gives the above multiple correlation formula (Formula (157)), but does not give the derivation.

for the multiple correlation coefficient as tabulated in the body of Table I. If we add another test or variable, n becomes 6 and we obtain the results given in Table II. In Table III, it is assumed that n, the number of "independent" variables, is 10.

TABLE I

$n = 5$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.50	.45	.42
$r_0 = .40$	.67	.60	.55
$r_0 = .50$	.83	.75	.69

TABLE II

$n = 6$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.52	.46	.42
$r_0 = .40$	.69	.62	.57
$r_0 = .50$	.87	.77	.71

TABLE III

$n = 10$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.57	.49	.44
$r_0 = .40$	.76	.66	.59
$r_0 = .50$	.94	.82	.74

From a comparison of Tables I and II, we see quite clearly that with only one exception ($r_0 = .50$; $r_i = .20$), the addition of another test has not increased the size of the multiple correlation coefficient by more than .02. Hence, for practical predictive purposes, we should probably be interested in five rather than six "independent" variables. In fact, even with ten "independent" variables, only one of the coefficients shows more than a .10 rise, and the construction of five more tests with the characteristics of those mentioned above would appear to involve an amount of work all out of proportion

to the resulting increase in the size of the multiple correlation coefficient. Hence we shall consider Table I only.

Since we would scarcely expect to obtain correlations with our criterion as low as or lower than the intercorrelations of our "independent" variables (which have been expressly selected to measure *different* aspects of the criterion), we may neglect the three such correlations (.45, .42, and .55). Since, also, it is too much to hope that the tests will combine to give us the highest expected correlations with the criterion as well as the lowest expected intercorrelations, we may likewise neglect the multiple correlation of .83.

As the result of this analysis of the three tables (which may be constructed and checked in less than an hour) we may conclude: (1) That more than five tests are not desirable, even if possible; (2) that the best result we may hope to get is represented by $R = .75$; and (3) that the worst result we may expect is represented by $R = .50$. If a correlation between these limits is satisfactory, we may well continue the investigation; if not, we must either find better tests or abandon the undertaking.

TABLE IV

	0	1	2	3	4	5
0	1	.4017	.6003	.2379	.6807	.3553
1	.4017	1	.2332	.1986	.2569	.1064
2	.6003	.2332	1	.1747	.4520	.2139
3	.2379	.1986	.1747	1	.2628	.0033
4	.6807	.2569	.4520	.2628	1	.2989
5	.3553	.1064	.2139	.0033	.2989	1

An example of the use of the formula in roughly approximating a multiple correlation coefficient when all the intercorrelations are known may be afforded by the use of the data in Table IV. Letting $r_0 = .4552$ (the average of the criterion correlations) and $r_i = .2201$ (the average of all the other correlations) we readily obtain $R_{0.12345} = .742$. Solution of the actual multiple correlation by determinants[1] yields $R_{0.12345} = .790$, indicating an error of $-.048$ in the approximation. The writer has approximated eight eight-variable multiple correlations by this method and found errors of $-.012$, $-.020$, $-.054$, $-.016$,

[1] Kelley, Truman L.: "Statistical Method." P. 309.

for the multiple correlation coefficient as tabulated in the body of Table I. If we add another test or variable, n becomes 6 and we obtain the results given in Table II. In Table III, it is assumed that n, the number of "independent" variables, is 10.

TABLE I

$n = 5$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.50	.45	.42
$r_0 = .40$	.67	.60	.55
$r_0 = .50$	.83	.75	.69

TABLE II

$n = 6$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.52	.46	.42
$r_0 = .40$	.69	.62	.57
$r_0 = .50$	.87	.77	.71

TABLE III

$n = 10$	$r_i = .20$	$r_i = .30$	$r_i = .40$
$r_0 = .30$	.57	.49	.44
$r_0 = .40$	.76	.66	.59
$r_0 = .50$	.94	.82	.74

From a comparison of Tables I and II, we see quite clearly that with only one exception ($r_0 = .50$; $r_i = .20$), the addition of another test has not increased the size of the multiple correlation coefficient by more than .02. Hence, for practical predictive purposes, we should probably be interested in five rather than six "independent" variables. In fact, even with ten "independent" variables, only one of the coefficients shows more than a .10 rise, and the construction of five more tests with the characteristics of those mentioned above would appear to involve an amount of work all out of proportion

to the resulting increase in the size of the multiple correlation coefficient. Hence we shall consider Table I only.

Since we would scarcely expect to obtain correlations with our criterion as low as or lower than the intercorrelations of our "independent" variables (which have been expressly selected to measure *different* aspects of the criterion), we may neglect the three such correlations (.45, .42, and .55). Since, also, it is too much to hope that the tests will combine to give us the highest expected correlations with the criterion as well as the lowest expected intercorrelations, we may likewise neglect the multiple correlation of .83.

As the result of this analysis of the three tables (which may be constructed and checked in less than an hour) we may conclude: (1) That more than five tests are not desirable, even if possible; (2) that the best result we may hope to get is represented by $R = .75$; and (3) that the worst result we may expect is represented by $R = .50$. If a correlation between these limits is satisfactory, we may well continue the investigation; if not, we must either find better tests or abandon the undertaking.

TABLE IV

	0	1	2	3	4	5
0	1	.4017	.6003	.2379	.6807	.3553
1	.4017	1	.2332	.1986	.2569	.1064
2	.6003	.2332	1	.1747	.4520	.2139
3	.2379	.1986	.1747	1	.2628	.0033
4	.6807	.2569	.4520	.2628	1	.2989
5	.3553	.1064	.2139	.0033	.2989	1

An example of the use of the formula in roughly approximating a multiple correlation coefficient when all the intercorrelations are known may be afforded by the use of the data in Table IV. Letting $r_0 = .4552$ (the average of the criterion correlations) and $r_i = .2201$ (the average of all the other correlations) we readily obtain $R_{0.12345} = .742$. Solution of the actual multiple correlation by determinants[1] yields $R_{0.12345} = .790$, indicating an error of $-.048$ in the approximation. The writer has approximated eight eight-variable multiple correlations by this method and found errors of $-.012$, $-.020$, $-.054$, $-.016$,

[1] Kelley, Truman L.: "Statistical Method." P. 309.

−.016, −.015, −.042 and +.001.[1] With the other problems on which this method has been tried (three to nine variables), errors of −.090, −.034, −.011, −.010, −.007, .000, +.001, +.002, and +.006 were obtained. In case any of the correlations between the criterion and an "independent" variable are negative, they should be made positive before averaging. This is allowable if for each such variable, say X_t, we substitute its negative X_t' and reverse the sign of every correlation involving variable t. Thus, the series $r_{01} = .52$; $r_{02} = .41$; $r_{03} = -.14$; $r_{12} = .49$; $r_{13} = .23$; $r_{23} = .25$ should be changed to $r_{01} = .52$; $r_{02} = .41$; $r_{03} = +.14$; $r_{12} = .49$; $r_{13} = -.23$; $r_{23} = -.25$ before the estimation of $R_{0.123}$.

As an example of the purely theoretical use of the formula, let us consider the special case in which the "independent" variables are mathematically independent ($r_i = .00$). The formula then becomes

$$R_{0.123\ldots n} = \sqrt{nr_0^2} = |r_0\sqrt{n}|$$

Table V gives the values of $R_{0.123\ldots n}$ which would result from various combinations of n and r_0. Immediately one asks the meaning of the entries greater than 1.00. The explanation is obvious. All such correlations result from assumed and fictitious data which it is utterly impossible ever to secure. It may be interesting to note that with ten mathematically independent tests criterion correlations even as low as .30 will allow excellent prediction. It may also be noted that four mathematically independent tests or measures will completely determine a fifth measure when $r_0 = .50$.

Two statements respecting the limitations of the formula should be made. Whenever the value of $R_{0.123\ldots n}$ is greater than 1.00 (*cf.* Table V) or imaginary (as when $r_0 \neq .00$; $r_i = -.10$; and $n > 11$), the assumed correlations are impossible and can never be obtained in real problems. Whenever the criterion correlations are unequal and their average is used for r_0, the true multiple correlation coefficient may be expected to be larger than the obtained coefficient.

---

[1] The first seven of these correlations connected each Alpha test (except test seven) with all the Beta tests. The omission of the correlation with Alpha test seven was necessitated because of two inaccuracies in the doubly recorded correlation coefficients. The other multiple correlation was between the combined scale and seven of the Alpha tests. See "Memoirs of the National Academy of Sciences," 1921, Vol. XV, pp. 634–651.

TABLE V

$r_i = .00$	$n = 2$	$n = 4$	$n = 10$	$n = 20$	$n = 50$
$r_0 = .00$	.00	.00	.00	.00	.00
$r_0 = \pm.10$	.14	.20	.32	.45	.71
$r_0 = \pm.20$	.28	.40	.63	.89	1+
$r_0 = \pm.30$	.42	.60	.95	1+	1+
$r_0 = \pm.40$	.57	.80	1+	1+	1+
$r_0 = \pm.50$	.71	1.00	1+	1+	1+
$r_0 = \pm.60$	.85	1+	1+	1+	1+
$r_0 = \pm.70$	.99	1+	1+	1+	1+
$r_0 = \pm.80$	1+	1+	1+	1+	1+
$r_0 = \pm.90$	1+	1+	1+	1+	1+

A SHORT METHOD FOR COMPUTING THE CORRELATION BETWEEN INTERCHANGEABLE VARIABLES

FLORENCE L. GOODENOUGH

Institute of Child Welfare, University of Minnesota

In problems such as those dealing with the correlation between the measurements of like-sexed twins, where no objective criterion exists for determining which member of a given pair should be considered the x-variate and which the y-variate, the double-entry method is commonly used. This method involves entering each pair of measurements on the correlation chart twice, so that each measurement is considered first as the x-variate, then as the y-variate.

A brief examination of the product-moment formula will show that under these circumstances the making of the scatter-diagram becomes unnecessary. Since each measurement is to be entered both as an x-variate and as a y-variate, the two frequency-distributions will be identical and will consequently have the same mean and standard deviation. We may therefore substitute σ^2 (the standard deviation of the frequency distribution resulting when all the scores for both series of measurements are arranged in the form of a single series) for $\sigma x \sigma y$ in the denominator of the usual formula, which thus becomes

$$r = \frac{\Sigma xy/n}{\sigma^2}$$

And since xy is equal to yx the sum of the paired products may also be computed in terms of the products of the deviations of the respective members of the pair from the common mean. If the computation is made from an arbitrary origin instead of the true mean, the usual correction should be made in the numerator. But since this correction is identical for the two variables, $c_x c_y$ may be written simply as c^2. The corrected formula then becomes

$$r = \frac{\frac{\Sigma xy}{n} - c^2}{\sigma^2}$$

where Σxy equals the sum of the products of the deviations of the members of each pair from the arbitrary origin, n is the number of pairs, c^2 is the square of the difference between the arbitrary origin and the true mean, and σ^2 is the square of the standard deviation of the single combined distribution.

A formula for correlating interchangeable variables without the use of the scatter diagram has previously been published by Furfey.[1] The method here described will yield the same results with a considerable shortening of the labor.

[1] Furfey, Paul H. A Formula for Correlating Interchangeable Variables. *Journal of Educational Psychology*, Vol. XVIII, 1927, pp. 122–124.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

Studies in the Comparative Abilities of Whites and Negroes. Joseph Peterson and Lyle H. Lanier. Mental Measurement Monographs, Serial No. 5, Feb., 1929. "The outstanding result of our tests of adults is an enormous and reliable superiority of whites over negroes in all four group intelligence tests, this being true of the wholly non-verbal International rotator tests as well as of the Binet group, the Myers, and the Atkinson test." The subjects were also tested with the Seashore musical tests; a tapping test, a substitution test; a form board; the Stenquist tests; and the Downey Will-Temperament tests.

The Handicap of a Limited Vocabulary. Frances E. Andrews. Chicago Schools Journal, Dec., 1928, 128–129. This investigation indicated a significant group of freshman high school pupils of average intelligence but of below average vocabulary due to home environment.

Measurement of Achievement

Sex Differences in History Retention. Sarah Janet Bassett. School and Society, Mar. 23, 1929, 397–398. With one hundred ninety cases, the boys proved slightly superior in total history retention, and especially superior in items relating to war or to some geographical place; the girls proved superior in history treating domestic conditions and home life.

Criteria for Judging Thought Content in Written English. Zelma Langdon Huxtable. Journal of Educational Research, Mar., 1929, 188–195. Compositions (N, 1200) were selected from Grades VII to IX. Five levels of thought complexity were differentiated by teacher judgment: inarticulate thoughts, unrelated thoughts, related thoughts reflective thoughts, creative thoughts.

Psychology of Learning and School Subjects

The Relation of Intelligence and of Mechanical Speeds to the Various Stages of Learning. W. R. Atkinson. Journal of Experimental Psychology, Apr., 1929, 89–112. Thirty subjects were tested with the Army Alpha and the Columbia Intelligence Examination, and then practised in card-sorting, number-code, number-alphabet. "Intelligence and learning ability, as measured by the tests used in this study, are not identical."

Certain Factors Underlying the Acquisition of Motor Skill by Pre-school Children. Florence L. Goodenough and Clara R. Brian. Journal of Experimental Psychology, Apr., 1929, 127–155. Twenty four-year-old children were practised in a ring-toss game. Three learning groups of equivalent ability were formed: (1) the uncoached group, (2) the suggested procedure group, and (3) the required procedure group. The third group made the greatest progress; and the second, the next most improvement. Individual learning curves are presented.

A Study of Erroneous Word Concepts in Reading. Frances R. Dearborn. The Elementary English Review, Jan., 1929, 3–6. A study of one hundred twenty-four word difficulties experienced by third and fourth grade children seemed to indicate these sources of error: (1) lack of experience with the word, thirty-nine per cent; (2) lack of word recognition or spelling knowledge, twenty-seven per cent; (3) mispronunciation, nineteen per cent; (4) mixed concepts previously formed, thirteen per cent; (5) like sounds of words that differ in spelling and meaning, six per cent; and (6) similar spelling, but different meaning, three per cent.

The Development of Independence in Word Recognition. Arthur I. Gates. The Elementary English Review, Apr., 1928, 113–115.

Remedial Work in Reading. The Elementary English Review, (1) Apr., 1928, 122–126; (II) May, 1928, 150–153. A short list of effective devices is included in the article.

Summary of Reading Investigations (July 1, 1927, to June 30, 1928) II. William Scott Gray, The Elementary School Journal, Mar., 1929, 496–509.

A Simple Technique for Determining Whether Children Knows the Meaning of Spelling Words. Carleton Washburne and Mable Vogel Morphett. Journal of Educational Research, Mar., 1929, 196–202. The authors found that words unfamiliar to children can be determined by the simple expedient of asking children to raise their hands while the list of words is being read.

Possible Value of Four Spelling Rules. Harry V. Masters. The Elementary English Review, Sept., 1928, 212–219. The author found that the four most common spelling rules applied to only fifty-one of the two hundred sixty-eight most difficult of the common words encountered by eighth grade pupils.

Statistical Methods

Accuracy in Calculation. Karl J. Holzinger. The Elementary School Journal, Mar., 1929, 510–517. The writer discusses the problem of significant figures and the rounding of numbers.

The Reliability of Rating Schemes. George R. Moon. School and Society, Mar. 16, 1929, 366–368. Teachers' ratings are compared with scholastic success, the two proving to be closely associated.

Curriculum

Children's Choices in Poetry as Affected by Teachers' Choices. Alice B. Coast. The Elementary English Review, May, 1928, 145–147. This study indicates a considerable overlapping between the choices of children and the choices of the teacher in a given class.

A Comparative Study in Children's Magazine Interests. Ruth Elizabeth Norris. The Elementary English Review, Oct., 1928, 241–245.

The Effect of Library Reading. Maude Vandiver. Educational Administration and Supervision, Mar., 1929, 190–201. An association was found between the pupils receiving reading certificates and the number of books read per year.

A Research Problem in Children's Reading. The Elementary English Review, Jan., 1929, 1–2. A study of the amount and the kinds of reading done by five hundred children in a public library.

Children's Choices of Poems. The Elementary English Review, June, 1928, 182–185. With groups of one hundred fifty to two hundred children, Grades I to III, poems by Edward Lear, A. A. Milne, and Walter De la Mare prove favorites.

What is Good Fraction Drill? Ross H. Beall. The Elementary English Review, Feb., 1929, 49–52. A study showing the distribution of the practice in the addition of common fractions and mixed numbers provided by three series of arithmetics for Grades IV to VIII.

Promise and Performance in Arithmetic Drill Service. Guy M. Wilson. Journal of Educational Research, Mar., 1929, 169–176. The writer suggests that in planning arithmetic drill, the adding of a "thought of" number to a "seen" number in column addition and of a "thought of" number to a "thought of" number in multiplication must be taken into account as the basic facts for the mastery of addition.

Language Text Books: A Study of Five Recent Seventh Grade Texts. Mildred Dawson. The Elementary English Review, (I) Feb., 1929, 43–46; (II) Mar., 1929, 69–73. A study of methods used and topics included in the various texts.

Administration and Supervision

The Principal and Progressive Movements in Education. Research Bulletin of the National Education Association, Mar., 1929. A summary of some recent studies regarding health, mental life, individual differences, character education, preschool education, and parent education.

Personality Traits of School Administrators. E. S. Lide. Educational Research Bulletin, Apr. 3, 1929, 141–143. The ranking by twenty-five judges of twenty-six traits is presented. The four most important of these traits were good judgment. breadth of interest, leadership, broadmindedness.

Parents' and Pupils' Report Cards. E. J. Ashbaugh. Educational research Bulletin, Apr. 3, 1929, 143–146. By a questionnaire the investigator determined the items of information which parents desire.

A Comparison of the Time Spent by First-year Students and Expected by Teachers in a State Normal School. N. A. Wade. Journal of Educational Research, Mar., 1929, 183–187. The average first-year student (*N*, 80) spends twenty-one hours and fifteen minutes per week in study; the average teacher expects him to spend twenty-three hours and twenty-nine minutes.

A Comparison of Teacher-training Centers with Schools Having a Population of Like Ability. Allan J. Williams. Educational Administration and Supervision, Mar., 1929, 184–189. With large numbers of pupils in each of two groups, the differences in achievement between teacher-training schools and others were insignificant.

How Shall We Section Beginning Foreign Language Classes? James Burton Tharp. The Modern Language Journal, Mar., 1929, 433–449. The experimental groups sectioned on basis of aptitude tests proved slightly superior to the control or unsectioned groups.

What Professional Magazines Do Prospective Educators Read? Walter Crosby Eells. School and Society, Apr. 6, 1929, 446–448. Articles were chosen most frequently from School and Society, School Review, a state journal, Educational Review, and Journal of the National Education Association.

Measurement of Special Abilities

Ocular Dominance Demonstrated by Unconscious Sighting. Walter Miles. Journal of Experimental Psychology, Apr., 1929, 113–126. A simple device (including photography) for determining the predominant eye is suggested. Results show high consistency, about sixty-five per cent of the groups tested proving "right-eyed."

Handedness. Ira M. Gast. The Journal of Educational Sociology, Apr., 1929, 487–491. A study of three hundred thirty-three individuals, twelve years or older, in 24 different activities, revealed that seventy per cent of the activities were performed by the right hand; eighteen per cent by the left; and twelve per cent by either hand.

Study Habits and Methods

How to Study in High School. E. E. Keener. Chicago Schools Journal, Jan., 1929, 167–169. Improvement appeared in those groups given remedial treatment in reading.

A Ten-minute Observation in the Library. Dennis C. Troth. School and Society, Mar. 9, 1929, 336–338. An interesting report of the time spent in conversation (twenty-nine per cent of total), in aimlessly looking around (15.5 per cent); in aimlessly leafing books (12.4 per cent), etc. by one hundred undergraduate students.

Miscellaneous

American Novelists Ranked. A Psychological Study. John M. Stalnaker and Fred Eggan. The English Journal, Apr., 1929, 295–307. A selected group of sixty-five outstanding critics rated the work of seventy-two contemporary American novelists of some importance and standing. The rating was an "estimate of their general literary merit as evidenced by their novels" indicated on a scale of ten divisions. In Group 1 (highest) appeared Willa Cather and Edith Wharton; in Group 2, Theodore Dreiser, James Branch Cabell, Sherwood Anderson, and Sinclair Lewis; in Group 10 (lowest), Harold Bell Wright, Thomas Dixon, Justin Strum, and E. R. Burroughs.

A Study in the Origins of Distinguished Living Americans. Roy Hinman Holmes. The American Journal of Sociology, Jan., 1929, 670–685. A study of origin, urban and rural, of men in Who's Who indicates a diminishing productivity from urban centers.

The Interrelation of Emotion and Intelligence. David L. Mackaye. The American Journal of Sociology, Nov., 1928, 451–464. A report of a number of case studies tending to show that an emotional upset also influences intelligence test and achievement test performances.

Hygienic Standards in Type and Format of Reading Materials. J. Herbert Blackhurst. The Elementary English Review, Apr., 1928, 101–103. A concise report of a few important studies regarding the hygiene of reading materials.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

The Value of Homogeneous Grouping, by T. Luther Purdom. Baltimore; Warwick and York, 1929. Pp. 99.

The study reported in this first volume of the new "University Research Monographs" was made for the purpose of investigating "the value of homogeneous grouping made on the basis of intelligence tests" and "carried out under normal prevailing conditions in the different schools of the state." The title chosen is unfortunately over-comprehensive, for the experiment deals only with first year high school pupils, and the work was confined to classes in English and algebra. This, however, need not concern us too much, for we all are prone to be over-ambitious in our choice of title.

After a chapter in which the previous experiments are reviewed, the second chapter is devoted to the method of attacking the problem. At the beginning of the semester in 1923 the Terman Group Intelligence Test was given, and equivalent experimental and control groups were formed. In the smaller schools the experimental groups were divided into two sections, the lower and higher halves on the basis of the test scores, and in the larger schools a greater number of sections was possible. The control groups were necessarily divided into classes, but care was taken that these should contain pupils of all ranges of ability. Sex and chronological age were also considered in making up the various groups in order that these factors should not influence the results.

All pupils were also given at the beginning of the semester four standardized tests in English: Brigg's English Form Test, Van Wagenen's English Literature Scale, and two parts of Kirby's Grammar Test. No tests were given in alegbra as this subject had not been studied by any of the pupils.

Each teacher taught both control and experimental groups, and thus another factor was held constant. No special methods were suggested for use in dealing with the different groups, but each teacher

was instructed to cover as much ground as possible and to teach her pupils all she could of her subject.

At the end of the semester the English tests were repeated, and the Hotz Algebra Tests were given. The teacher's marks and opinions were also studied for evidence as to the value of homogeneous grouping.

The findings at which Purdom arrives are briefly as follows: The pupils in homogeneous sections did not gain more in standard test results than those in heterogeneous sections; the semester grades were higher in algebra but lower in English; the homogeneous sections did not cover more material; no intelligence group profited by the grouping; the failures were not reduced by homogeneous grouping; the teacher's opinion, although somewhat divided was, on the whole, favorable to the plan.

The real meaning of the experiment, however, goes beyond this particular set of findings, although unfortunately Purdom does not appear to be aware of this fact, and certainly makes no mention of it. The conclusions which Purdom sets down should be preceded by some such statement as the following, set in black face type, "*When homogeneous grouping is made on the basis of the results of a single intelligence test unmodified by other data, and when no provision is made for difference in methods and materials to be used in handling the groups at different ability levels . . .*"

Here, I believe, lies the crux of the matter. A survey of the literature brings out very clearly that if ability grouping is to produce the benefits which are claimed for it the grouping must be very carefully made, and after the groups are established they must be taught by methods which are suited to their respective abilities. This latter problem should be carefully considered with the teachers before work is done, and no teacher should undertake to teach a special group until she understands what she is to do and how she is to do it. One junior high school principal who is getting very satisfactory results spent a whole year studying the plan with her teachers before she attempted grouping.

Quite as important is the fact that no single measure has proved sufficiently reliable to be used alone as a basis of grouping. Possibly it is not necessary to take into account so much as does the elaborate plan of Ryan and Crecilius, but it is quite certain that many mistakes in grouping will be made if the scores of a single test are used as the sole basis for classification.

The real import of the experiment, then, lies in the demonstration of the fact that the mere putting together of pupils who make like scores on a single intelligence test and letting them be taught in the ordinary way will not produce any important change in their achievement. In other words, form of organization alone is not enough. The advantages and desirable outcomes of ability grouping can only be realized when groups are made up with the utmost care, when courses of study, methods and materials are differentiated for the different ability levels, when the teachers of the various groups understand and practice the techniques appropriate to these groups, and when both teacher and supervisory officers thoroughly understand and appreciate the philosophy underlying this form of school organization.

EDWARD A. LINCOLN.
Graduate School of Education, Harvard University.

The Fundamentals of Human Motivation, by Leonard T. Troland. New York: D. Van Nostrand Co., Inc. Pp. XIV + 521.

The present book is intended to be a systematic treatment of the facts and problems of human motivation. The author states that he is not aware of the existence of any comprehensive treatment which presents all the pertinent facts and theories of what may be termed orectic psychology. We agree with him in this but are inclined, after reading his book, to wonder whether the time is ripe for such a comprehensive treatment.

The book opens with the usual definitions and historical review, then goes on to discuss the nature of neuromuscular responses, reflexes, instincts, and the like. These are discussed in a somewhat critical manner and lead about half way through the book to the development of a theory of Nociception, Beneception, and retroflex action. The remainder of the book is taken up with further description and application of this theory.

Troland owes much of his inspiration to McDougall and Morton Prince. He occupies much space in description of reflexes, simple and complex, and their relation to instincts. He makes out a list of instincts which disagrees with those of other people in certain respects. Nevertheless he defines an instinct much in the same way as others. Why is it that everyone who writes on instincts makes up a new list? Is it not because instincts do not really exist as functional units, but are only loose descriptive terms used in classifying behaviour?

Developing out of this physiological discussion, he describes his new terms, Nociception and Beneception, as follows: "Nociception is accompanied by a decreasing of the conductance of operating cortical adjustors, whereas Beneception is accompanied by an increasing of the conductance of operating cortical adjustors. We may assign to the process that is thus described the name retroflex action." Out of this somewhat weird proposition he develops his theory of human motivation. It seems to be nothing but an attempted physiological explanation of the old hedonistic principle, modernized by giving it biological utility. Troland realizes the necessity of relating this in some way to a doubtful consciousness, but does not manage to explain the problem any better than McDougall did with his psychophysical parallelism.

Perusal of the book merely serves to bring out more prominently than ever how little is known other than by metaphysical speculation (if that can be called knowing) of human motivation. There is scant reference to experimental work, this being due, of course, to the paucity of such work. The most promising attempts in experimental research were the experiments conducted by John B. Watson. At the beginning he apparently adopted a questioning attitude almost saying, "The origins of human conduct have never been satisfactorily investigated. Philosophers and pseudopsychologists have theorized for centuries about it. It will make no theory but will find out." His finding out was very unfortunately stopped after he had done a small amount of experimental work, and obtained results which ought never to be regarded as anything but suggestive. He has since contented himself with theorizing like James, McDougall, Morton Prince, Allport, Kohler, Graham Wallace, and others. Troland adds one more. When some group of persons seriously take up the work which Watson dropped and when unequivocal results are obtained the time may be ripe for volumes on human motivation of scientific value. The present book presenting opinions approached in a scientific spirit, does not constitute a satisfactory substitute.

C. S. Slocombe.

Boston Elevated Railway.

Introduction to Education, by F. L. Clapp, W. L. Chase and Curtis Merriman. Boston: Ginn and Co., 1929. Pp. 569.

This "Introduction to Education" is a textbook for the elementary survey course in education which really surveys and really integrates. Most publications under such titles as "Introduction to Teaching" are

handicapped by too limited a point of view on one hand or by a compartmented organization on the other. Clapp, Chase and Merriman succeed remarkably in combining an encyclopædic inclusiveness with reasonable continuity.

The presentation of material is essentially concrete and inductive. Tables, charts and case illustrations are numerous, yet are included to substantiate and clarify general notions rather than for their own sakes. This approach makes the text its own source book, and therefore both teachable and adaptable to the individual instructor's organization of the course.

One unique and valuable approach is in the consideration of European school systems and educational practices. Not only is this worth while for its own sake, but it furnishes a background for the comparison and criticism of the familiar features of American education. The concepts of elementary and secondary education, for example, are greatly enriched by the realization of their European meanings. Foreign organization and curricula serve to jar the student from the set complaisancy with which he is likely to regard the school system in which he was reared.

Although the work is chronologically up to date, and the point of view is in general progressive, there is one omission in "Introduction of Education" which is slightly painful. No mention is made of the "child-centered" elementary school, nor of the most recent aspects of the revision of the curriculum. Perhaps this omission is wise in a textbook, due to widely divergent points of view that instructors are likely to have on these subjects. This field is the only one, however, in which I have felt it necessary to assign other readings in connection with the book, as used in an "Introduction to Teaching" course.

This volume seems, however, the outstanding recent work of its kind. It will be widely used by those who feel that a factual, meaty, comprehensive survey course is a valuable foundation for the study of education.

LAURANCE F. SHAFFER.

Carnegie Institute of Technology.

Better Schools, by Carlton Washburne and Myron M. Stearns. New York: The John Day Co., 1928. Pp. XVI + 342.

In the matter of book reviews enthusiasm for a particular *opus* is often taken as a mark of deficiency of some sort on the part of the

reviewer. There are other reasons, moreover, why reviewers should hedge or openly throw stones; yet one can hardly be accused of astigmatism for the sake of reviewing educational texts. This "Survey of Progressive Education in American Public Schools" deserves praise. It meets a need. With so many new techniques, methods of organiza tion, and researches being reported the professional practitioner feels the need of a periodic summation to become properly oriented in the field of American education today. How much more is it necessary that parents desirous of giving their children the advantage of the best should welcome such a survey.

The purpose of this book is given in this manner in the Introduction: "There is not . . . any non-technical book that selects for the layman and unspecialized teacher or school administrator some of the outstanding and successful experiments and presents them in a way that will help one to understand and profit by American new school ways." It is an ambitious purpose. The treatment is concise and sticks to the purpose. It does not seem possible that so much could be found between the covers of such a comparatively small volume.

A justification for progressive schools—schools with new aims and methods—is found in the precarious situation of western civilization. If it is to be preserved the next generation must be educated for the purpose. Then how are better schools obtained? A story as strange as fiction is that of the fight for better schools in Denver two decades past. Methods of enlisting the cooperation of communities and teachers are given ample discussion as are new methods of organization. But organization is not all of the answer for better schools. What shall be taught? This question introduces an interesting presentation of some of the outstanding efforts at reconstructing the curriculum, giving vocational guidance, and connecting the classroom with life. Barnes' work in the rural schools of St. Louis County, Minnesota; Rochester's fine junior high schools in spite of Regent's Examinations; Dean Schneider's experiments in the Engineering School at Cincinnati; these and many other struggles for improvement too numerous to mention make this a most readable book. Other experimenters have been interested in better ways to teach. There is a very clear discussion of the effort to replace discipline with interest. The two movements centering around interest are made clear enough for any layman looking for the answer. The efforts being made at Winnetka and through the Dalton Plan are elaborated and evaluated for the intended audience. Child guidance is made understandable

for non-specialists. Morrison's new principles are clearly stated in non-technical language.

The above enumeration does not do justice to the book. It must be read to be appreciated. The book carries an air of authority with it. Not all the leaders or new methods are tabulated. It is to the credit of the authors that they selected representative examples in both cases. The combination of authors is a happy one. Mr. Washburne brings an excellent reputation in educational affairs. Mr. Stearns had an unusual opportunity in visiting so many schools throughout the country—an enviable experience. "Better Schools" is inspiring. One closes it reluctantly. Wherever schoolmen or laymen are gathered together (and the meeting of both in the Parent-Teacher Association holds as well) for the purpose of improving their schools, here is a source of well-organized information concerning the best in progressive education throughout the United States today.

J. H. COLEMAN.

Graduate Student, Teachers College, Columbia.

The Learning of the One Hundred Addition Combinations and the One Hundred Subtraction Combinations, by Frederic B. Knight, and Minnie S. Behrens. New York: Longmans, Green and Co., 1928. Pp. XIII + 82.

Educational procedure has long been handicapped by a lack of knowledge of some very simple things. Until recently we have been without reliable information concerning the amount of practice required to master and remember the correct answer to each of the fundamental combinations. Now this information is available through the careful studies of Dr. Knight and Miss Behrens.

The studies were carefully conducted and based upon actual learning conditions. The experimental group consisted of twenty-five second grade children for addition and fifteen children for subtraction. Exact records were kept of each child's work on each combination. The records include all actual responses, the time of all responses in tenths of seconds, the place and nature of all instruction, and the occurrence and nature of each error. The time records were obtained by the use of a simple kymograph.

Records were obtained from each child until a one hundred per cent response was obtained. Then tests were given once each week until the mastery of the combinations was so thorough that any failure could be considered as a decidedly "sport" affair. From the results

the authors are able to tell us how many repetitions were necessary in order to fix each combination in the mind of each child, thus giving a conclusive answer to the troublesome question of the relative difficulties of the several combinations. The results are summarized for both medium and slow pupils.

With reference to type errors the conclusions of the authors are quite different from those of Myers. Myers reports a strong tendency for the same error to persist but no such tendency is manifested in the present study.

The authors present much evidence to show that the relative difficulty of the combinations cannot be properly shown by a study of the errors made by upper grade children. This is of importance because our former knowledge of the relative difficulty of the combinations rests almost entirely on the basis of data obtained from such errors.

The monograph will be of notable value to all teachers and supervisors because it settles rather definitely the vexatious question of how much drill is needed on the several combinations in addition and subtraction. It is to be hoped that these authors may soon furnish us with equally valuable data concerning the difficulty of eaeh of the combinations in multiplication and division. W. J. Osburn.

Ohio State University.

Springs of Human Action, by M. K. Thomson. New York: D. Appleton and Co., 1927. Pp. XVI + 501.

The purpose of this book is to make a comprehensive study of motives, showing their relationship, practical application, and synthesis of simple motives into compound ones. In addition to this it is proposed to distinguish between unconscious and conscious, primary and secondary, innate and acquired drives. To accomplish this, the author discards the methods of treatment that have been used since he believes that no one method is adequate to the subject and chooses what he calls a comprehensive method which is for the most part a descriptive list.

The order of treatment is determined by a sequence from the "mechanical, innate, simple, and individualistic motives to the purposeful, complex, psychical, and social ones." The comprehensive study of motives begins with those due to physical activity and put forth by those who support dynamic psychology. Reflexes, glandular activity, habits, and instincts are then discussed. Emotions, feel-

ing, ideas, interest, will, choice, personality, self-regarding motives, and prejudice follow. The group of abnormal and suppressed motives in their relation to human conduct and lastly the motives of group suggestion, social motives, love, values, morale, economic, and æsthetic motives are described. Under these headings everything from tropisms to potential immortality is mentioned as a possible motive for conduct. With such an amount of material to be covered each topic is necessarily treated very briefly. A chapter on the interaction of motives, which seems significant in such a study, is given only two pages and a half at the end. A few motives are built up (the author prefers the synthetic rather than the analytic method) but there is no suggestion of how this may be done generally.

Obviously this is a thoroughly comprehensive study but it unfortunately gives more an encyclopediac and conventional effect than anything else. A broad systematic view is taken, but it is rather bewildering in effect. In fact the only purpose accomplished is a description of all the motives the different psychologists advocate, and by including all motives, the description of each is little more than a definition.

The author hopes to harmonize the various points of view of psychologists but the difficulties involved in such a unification are scarcely comprehended in this book. It is also intended for use as an introduction to any social science, as a text for the study of motivation, and as popular reading. Although interesting, his style will turn away the scientist by its catch phrases, its sweeping statements which are not very accurate, and by the superficiality of treatment. Its presupposition of a considerable knowledge not only in psychology but in other sciences as well together with its failure to make clear distinctions and show relationships render it unfit as a textbook and affect it even more as a book for popular reading. It is to be feared that these classes of readers are so diverse in their requirements that the same book could scarcely be expected to satisfy them all.

The value of a descriptive outline cannot, have the same value as an original and definite point of view on the subject, but any outline of a subject serves as a foundation for a study of the subject and so saves the labor of compiling such a table for one's own use. The book reads like a popular survey course for undergraduates, and as such may prove useful to the professor who prefers a descriptive approach to the subject rather than a study of the views of the different schools of psychology.

C. S. Slocombe.

Personnel Research Federation, N. Y. City.